Value-guided remapping of sensory cortex by lateral orbitofrontal cortex

Banerjee, Abhishek ; Parente, Giuseppe ; Teutsch, Jasper ; Lewis, Christopher ; Voigt, Fabian F ; Helmchen, Fritjof

Abstract: Adaptive behaviour crucially depends on flexible decision-making, which in mammals relies on the frontal cortex, specifically the orbitofrontal cortex (OFC)[1−9]. How OFC encodes decision variables and instructs sensory areas to guide adaptive behaviour are key open questions. Here we developed a reversal learning task for head-fixed mice, monitored the activity of neurons of the lateral OFC using two-photon calcium imaging and investigated how OFC dynamically interacts with primary somatosensory cortex (S1). Mice learned to discriminate 'go' from 'no-go' tactile stimuli[10,11] and adapt their behaviour upon reversal of stimulus-reward contingency ('rule switch'). Imaging individual neurons longitudinally across all behavioural phases revealed a distinct engagement of S1 and lateral OFC, with S1 neural activity reflecting initial task learning, whereas lateral OFC neurons responded saliently and transiently to the rule switch. We identified direct long-range projections from lateral OFC to S1 that can feed this activity back to S1 as value prediction error. This top-down signal updated sensory representations in S1 by functionally remapping responses in a subpopulation of neurons that was sensitive to reward history. Functional remapping crucially depended on top-down feedback as chemogenetic silencing of lateral OFC neurons disrupted reversal learning, as well as plasticity in S1. The dynamic interaction of lateral OFC with sensory cortex thus implements computations critical for value prediction that are history dependent and error based, providing plasticity essential for flexible decision-making.

DOI: https://doi.org/10.1038/s41586-020-2704-z

Letters

Value-guided remapping of sensory cortex by lateral orbitofrontal cortex

Abhishek Banerjee[1,3†], Giuseppe Parente[1*], Jasper Teutsch[1,3*], Christopher Lewis[1], Fabian F. Voigt[1,2], and Fritjof Helmchen[1,2†]

[1]Laboratory of Neural Circuit Dynamics
Brain Research Institute, University of Zürich
[2]Neuroscience Center Zürich
Winterthurerstrasse 190, 8057 Zürich, Switzerland

*These authors have contributed equally to this work
[3]Current Address: Biosciences Institute, Newcastle University, United Kingdom

Number of main text pages: 29
Number of main figures: 4
Total number of words (≤1500): 1729
Summary: (<300) 216
Number of references (≤30): 30

†Correspondence:

Abhishek Banerjee
E-mail: abhi.banerjee@newcastle.ac.uk
Phone: +44 191 208 52 27
and
Fritjof Helmchen
E-mail: helmchen@hifo.uzh.ch
Phone: +41 44 635 33 02

Keywords: Orbitofrontal cortex, Sensory processing, Reversal learning, S1, Reward-history, Credit assignment, Rule-switch

Adaptive behaviour crucially depends on flexible decision-making, which in mammals relies on frontal cortex, specifically the orbitofrontal cortex (OFC)[1-9]. How OFC encodes decision variables and instructs sensory areas to guide adaptive behaviour are key open questions. Here we developed a reversal learning task for head-fixed mice, monitored the activity of lateral OFC neurons using two-photon calcium imaging, and investigated how OFC dynamically interacts with primary somatosensory cortex (S1). Mice learned to discriminate go/no-go tactile stimuli[10,11] and adapt their behaviour upon reversal of stimulus–reward contingency ('rule-switch'). Imaging individual neurons longitudinally across all behavioural phases revealed a distinct engagement of S1 and lateral OFC: whereas S1 neural activity reflected initial task learning, lateral OFC neurons saliently and transiently responded to the rule-switch. We identified direct long-range projections from lateral OFC to S1 that can feed this activity as value prediction error back to S1. This top-down signal updated sensory representations in S1 by functionally remapping responses in a small outcome-selective neuronal subpopulation that was also sensitive to reward history. Functional remapping crucially depended on top-down feedback as chemogenetic silencing of lateral OFC neurons disrupted reversal learning as well as plastic changes in outcome-selective S1 neurons. The dynamic interaction of lateral OFC with sensory cortex thus implements history-dependent, value prediction error-based computations and plasticity essential for flexible decision-making.

Main Text

Animals adapt their behaviour to variable contextual changes in the environment. Central to adaptive behaviour is value-guided decision making, the ability to flexibly associate stimuli with preferred actions based on reward-history. Deficits in behavioural flexibility characterise brain disorders such as autism and schizophrenia[1]. In mammals, the prefrontal cortex is the locus of value-guided decision-making[2,3], with the OFC implicated in cognitive evaluation of stimulus-outcome associations[4–7]. OFC is a higher-order area with extensive connections to sensory cortices and subcortical structures of the reward system[8,9]. However, how neurons in OFC respond to changing reward contingencies is poorly understood. Further, whether OFC neurons instruct sensory areas to remap stimulus-outcome associations in support of adaptive behaviour is unclear.

To study flexible decision-making, we employed a reversal learning paradigm based on tactile discrimination. We trained mice to perform a 'go/no-go' texture-discrimination task[10] (**Fig. 1a**; P100 vs. P1200 sandpaper as go vs no-go texture; **Methods**). Once task performance reached expert level (discriminability index *d'* > 1.5), we implemented a 'rule-switch' by reversing the stimulus-reward contingency (**Fig. 1b**). Mice reached high *d'* values during initial learning ('learning naïve', *LN*, through 'learning expert', *LE*), decreased performance after reversal, and re-learned the task ('reversal naïve', *RN*, through 'reversal expert', *RE*) (**Fig. 1c**, **Extended Data Fig. 1**, n = 11 mice). Reversal learning was significantly faster, and performance remained stable over weeks (**Fig. 1c**, **Extended Data Fig. 1**). Task performance depended on sensory input and was independent of initial go-texture (n = 2 mice; **Extended Data Fig. 1**). Mice developed anticipatory whisking and well-timed licking during initial learning[11]. Following the rule-switch, the overall whisking behaviour did not change but mice transiently reverted to delayed licking before re-learning (*RE*, **Extended Data Fig. 2**). We investigated two areas implicated in task-learning: barrel cortex in the S1, important for tactile discrimination and sensory learning[12], and the lateral OFC (lOFC) that is critical for outcome-value assignment[8]. To examine the necessity of these areas, we expressed inhibitory DREADD receptors (hM4Di) in excitatory neurons in either S1 or lOFC (histology and electrophysiological validation in **Extended Data Figs. 3** and **4**). Inhibiting S1 neurons during initial training (via daily CNO injections before each behavioural training sessions during *LN* and *LE* periods) prevented task acquisition (**Fig. 1d**). Inhibiting neurons in lOFC, but not medial OFC[7], after the rule-switch (*RN* and *RE*) impaired reversal learning and increased perseverative errors (**Fig. 1d-f**, **Extended Data Fig. 3**). Interestingly, lOFC-silenced mice could still learn a new stimulus-outcome association (a new rewarded texture P600; **Fig. 1f**). Overall, these results indicate a dissociation of learning and reversal learning involving S1 and lOFC, respectively.

To monitor neuronal activity in lOFC and S1 during learning and reversal learning, we performed *in vivo* two-photon Ca^{2+} imaging in transgenic mice expressing GCaMP6f in superficial layer (L)2/3 excitatory neurons. We imaged lOFC, located deep in frontal cortex[13,14], via a gradient-index lens placed in a chronically implanted cannula (**Fig. 2a**; **Extended Data Fig. 5**; **Methods**, n = 4 mice). Cannula-implanted mice showed no whisking or behavioural impairments (**Extended Data Fig. 5**). We observed large Ca^{2+} transients in lOFC neurons particularly during the reward-outcome window (**Fig. 2a**). A longitudinally measured example neuron displayed modest reward-related activity during initial learning (*LE*), but large and robust responses to unexpected rewards immediately after the rule-switch (*RN*) (**Fig.**

2b). This activity was transient (*RN*) and decreased as mice re-learned the task (*RE*). Averaging across all lOFC neurons revealed the same pattern: a significant increase in the amplitude of reward-related Ca^{2+} transients after the rule-switch (*LE→RN;* **Fig. 2c**). These findings are consistent with lOFC encoding deviations from expected outcome-value following rule-switch[15]. In agreement with this, the response of lOFC neurons to a third rewarded texture (P600), associated with a constant small reward unaffected by reversal, remained unchanged (**Extended Data Fig. 6**). In contrast, L2/3 neurons in S1, imaged through a chronic cranial window (n = 5 mice), exhibited Ca^{2+} transients during stimulus-presentation and reward-outcome window (**Fig. 2d**). Responses to the rewarded go-texture emerged during learning (*LN→LE*), decreased following the rule-switch (*RN*), and were remapped to the new go-texture (*RE*) (an example neuron, **Fig. 2e**). Response remapping was significant across S1 L2/3 neurons (**Fig. 2f**). Response pattern was similar for anatomically identified S1→lOFC projection neurons (n = 3 mice; **Extended Data Fig. 7**). The dissociation was also evident in the fraction of active neurons in the periods of highest engagement: *LE* and *RN* for lOFC, versus *LE* and *RE* for S1 (**Fig. 2c** and **2f**).

Do neurons selective for rewarded hit trials retain selectivity for the old go-texture, or remap to the new go-texture after reversal, i.e. are they more stimulus- or outcome-selective? Longitudinal measurements of lOFC and S1 neurons permitted quantification of their response stability or flexibility upon rule-switch. To quantify response selectivity of active neurons, we defined an ROC-based hit/CR selectivity index (*SI,* ranging from -1 to 1, permutation test, $p < 0.05$; **Methods**; **Extended Data Fig. 8**)[16]. We focused on *SI* values for the reward-outcome window. Note that the *SI per se* cannot distinguish between stimulus- and outcome-selectivity because these trial-types differ in both texture-type and action-outcome. However, comparing *SI* values before and after the rule-switch reveals whether a neuron reverses (stimulus-selective) or maintains (outcome-selective) its *SI* sign. **Figure 3a** schematically presents the five major classes of *SI* changes and their distribution in a 2D before-after plot. Note that each neuron may have mixed stimulus- and outcome-selectivity (projections onto the diagonals). To assess both the immediate effect of the rule-switch and stable adaptation after re-learning, each neuron was classified to a major class twice (*LE→RN* and *LE→→RE*, respectively; **Fig. 3a**). Among 107 chronically imaged lOFC neurons (n = 3 mice), we found a preponderance of outcome-selective neurons that responded strongly to new-hit trials immediately following rule-switch (*RN*) (**Fig. 3b-c**). Additionally, some lOFC neurons lost or gained selectivity. This distribution persisted after re-learning (*LE→→RE*, **Fig. 3d; Extended Data Fig. 8**). In contrast, S1 neurons were more stimulus- than outcome-selective following reversal

(*LE→RN,* 18% of 218 neurons; n = 4 mice; **Fig. 3e-f**). However, the selectivity of S1 neurons changed markedly during re-learning (*LE→→RE*) with a large subpopulation functionally remapping to the new, rewarded go-texture (**Fig. 3g; Extended Data Fig. 8**). Moreover, a subpopulation of previously inactive or non-selective neurons acquired outcome-selectivity. Changes were similar for identified S1→lOFC projection neurons (**Extended Data Fig. 7**). An analogous analysis of texture-touch-evoked responses during stimulus-presentation likewise revealed an overall remapping towards the new go-texture (*RN→RE*, **Extended Data Fig. 9**). The link between functional subclasses and behavioural variables, especially reward-modulation of outcome-selective neurons was further confirmed by GLM[17] (**Extended Data Fig. 10; Methods**). These results suggest that lOFC neurons exhibit a value-guided response immediately following a rule-switch. In contrast, a subpopulation of S1 neurons initially retains the learned stimulus-value association and functionally remaps upon re-learning.

Is delayed S1 remapping causally dependent on lOFC? To investigate the existence of OFC→S1 projections in mice, we injected retrograde AAV-retro/2-tdTomato into L2/3 of S1. Whole-brain light-sheet microscopy[18] of cleared samples (n = 2) revealed dense S1-projecting neurons primarily in L2/3 and L5 of lOFC (**Fig. 4a**). Chemogenetic silencing of lOFC neurons after the rule-switch (*RN* through *RE*) impaired remapping of S1 neurons (**Fig. 4b**; **Extended Data Fig. 8**; n = 4 mice). The effect is best seen in the marginal distributions for the three salient learning periods. Unlike in control mice, a significant fraction of S1 neurons in lOFC-silenced animals preserved their selectivity, failing to remap during re-learning (cumulative distributions, two-sample Kolmogorov-Smirnov test) (**Fig. 4c**). Lateral OFC silencing also prevented *RN→RE* remapping of texture-touch-evoked responses (**Extended Data Fig. 9)**. We additionally tracked neuronal fate by comparing the assigned classes for *LE→RN* and *LE→→RE* transitions. Whereas a fraction of non-selective and lost-selectivity S1 neurons (*LE→RN*) normally gained selectivity for the new go-texture (*LE→→RE*), such recruitment did not occur in lOFC-silenced mice (**Extended Data Fig. 8; Methods**). These findings further confirm that S1 remapping crucially depend on top-down input from OFC.

Finally, we leveraged the sensitivity of lOFC neurons to reward-history to examine the mechanism by which lOFC influences S1 remapping. Most lOFC neurons that responded to new-hit trials also responded to FA trials immediately after reversal (*RN*), revealing lOFC neurons' sensitivity to deviations from expected outcome (**Fig. 4d-e**). We computed a 'reward-history modulation index' (RHMI) for lOFC and S1 neurons by comparing hit trials immediately preceded by a hit or a FA

(**Fig. 4f**; **Methods**). While outcome-selective neurons in lOFC exhibited significant reward-history-dependent response modulation before (*LE*) and after (*RN*) rule-switch, RHMI was significant in S1 for outcome-selective and acquired-selectivity neurons, but not other classes, after re-learning (*RE*). History-dependent modulation of S1 neurons was absent in lOFC-silenced mice indicating that lOFC is critical for the functional reorganisation of S1 (**Fig. 4g**; **Extended Data Fig. 10**). These findings corroborate the notion that lOFC encoding of outcome-value is essential to the functional remapping of S1 neurons in support of flexible decision-making.

Adaptive behaviour is shaped by sensory evidence and prediction of outcome-value of future choices. Predictions can shape perception[19] and OFC estimates the expected value of choices to achieve desirable outcomes, such as increased reward[20]. Our experiments revealed a crucial role of lOFC neurons in encoding prediction-error, partly resembling classical dopamine responses[21,22]. Critically, lOFC projections to S1 convey this teaching signal that drive remapping of sensory cortex (**Fig. 4h**). Tracking both positive and negative outcome-values, lOFC neurons may represent ongoing neural estimates of position on a value map[20]. Pharmacogenetic silencing revealed the necessity of lOFC to achieve flexibility as previously shown in rodents[23] (while OFC silencing showed mixed effects in non-human primates[6]). Outcome-value signals from lOFC are likely to interact via a rich assortment of projections to integrative cortical areas like the retrosplenial cortex[24], and subcortical structures including the basolateral amygdala[25] and the mediodorsal thalamus[26]. Further, we found that a small subpopulation of S1 neurons do not function as simple sensory feature detectors, but flexibly remap according to task context and reflect reward-history[17,27], characteristics expected in higher order areas, like OFC, but not in primary sensory areas. The cellular and circuit mechanisms enabling such remarkable plasticity remain to be determined, but may involve neuromodulators such as serotonin[28] or long-range, layer-specific excitatory and inhibitory interactions[29]. The existence of a reward-valence signal in the primary sensory cortex and its modulation by higher-order inputs has important implications for reinforcement learning algorithms[30]. Taken together, this study revealed local and long-range circuit interactions crucial to flexible sensory processing and adaptive decision-making.

Acknowledgements

This work is supported by a H2020 Marie Skłodowska-Curie fellowship (CIRCDYN, Grant number: 709288) and a NARSAD Young Investigator award (Grant number: 24941) from the Brain & Behavior Research Foundation to A.B., and grants from the Swiss National Science Foundation (Grant number: 310030B_170269), and the European Research Council (ERC Advanced Grant BRAINCOMPATH, Grant number: 670757) to F.H. We thank B. Grewe for showing us the preparation for GRIN lens imaging and M.E. Schwab for the use of equipment for open-field and ladder-rung test. We thank S. Carta, L. Shumanovski, D. Göckeritz, L. Egolf, C. Rickenbach for various assistance. We thank W. Senn, F. Lucantonio, M. Goard, M. Pignatelli, and B. Scholl for helpful discussions on the manuscript. The authors declare no conflict of interest or competing financial interest.

Author contributions

A.B. conceived the project. A.B. and F.H. designed the study. A.B and J.T. carried out all experiments except multi-unit electrophysiology experiments (performed and analysed by C.L.). A.B and G.P. analysed all data, except whisking kinematics data (analysed by C.L.). F.F.V developed light-sheet microscope and imaged cleared brains together with A.B.; A.B. and F.H. wrote the manuscript with comments from G.P. and C.L.

Data availability

The data that support the finding of this study are available upon reasonable request from the corresponding author.

References

1. Fettes, P., Schulze, L. & Downar, J. Cortico-striatal-thalamic loop circuits of the orbitofrontal cortex: Promising therapeutic targets in psychiatric illness. *Front. Syst. Neurosci.* **11**, 25 (2017).
2. Miller, E.K. The prefrontal cortex and cognitive control. *Nat. Rev. Neurosci.* **1**, 59–65 (2000).
3. Fuster, J.M. The prefrontal cortex-an update: Time is of the essence. *Neuron* **30**, 319–33 (2001).
4. Rolls, E.T. The orbitofrontal cortex and reward. *Cereb. Cortex* **10**, 284–94 (2000).
5. Izquierdo, A. Functional heterogeneity within rat orbitofrontal cortex in reward

learning and decision making. *J. Neurosci.* **37**, 10529–10540 (2017).
6. Rudebeck, P.H. & Murray, E.A. The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84**, 1143–1156 (2014).
7. Rushworth, M.F.S., Noonan, M.P., Boorman, E.D., Walton, M.E. & Behrens, T.E. Frontal cortex and reward-guided learning and decision-making. *Neuron* **70**, 1054–1069 (2011).
8. Wallis, J.D. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30**, 31–56 (2007).
9. Carlén, M. What constitutes the prefrontal cortex? *Science (80-.).* **358**, 478–482 (2017).
10. Chen, J.L., Carta, S., Soldado-Magraner, J., Schneider, B.L. & Helmchen, F. Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex. *Nature* **499**, 336–340 (2013).
11. Chen, J.L. et al. Pathway-specific reorganization of projection neurons in somatosensory cortex during learning. *Nat. Neurosci.* **18**, 1101–1108 (2015).
12. Petersen, C.C.H. Sensorimotor processing in the rodent barrel cortex. *Nat. Rev. Neurosci.* **9**, 533-546 (2019).
13. Bissonette, G.B., Schoenbaum, G., Roesch, M.R. & Powell, E.M. Interneurons are necessary for coordinated activity during reversal learning in orbitofrontal cortex. *Biol. Psychiatry* **77**, 454–464 (2015).
14. Jennings, J.H. et al. Interacting neural ensembles in orbitofrontal oortex for social and feeding behaviour. *Nature* **565**, 645–649 (2019).
15. Rushworth, M.F.S., Noonan, M.P., Boorman, E.D., Walton, M.E. & Behrens, T.E. Frontal cortex and reward-guided learning and decision-making. *Neuron* **70**, 1054–1069 (2011).
16. Pho, G.N., Goard, M.J., Woodson, J., Crawford, B. & Sur, M. Task-dependent representations of stimulus and choice in mouse parietal cortex. *Nat. Commun.* **9**, 2596 (2018).
17. Ramesh, R.N., Burgess, C.R., Sugden, A.U., Gyetvan, M. & Andermann, M.L. Intermingled ensembles in visual association cortex encode stimulus identity or predicted outcome. *Neuron* **100**, 900–915 (2018).
18. Voigt, F.F. et al. The mesoSPIM initiative: Open-source light-sheet microscopes for imaging cleared tissue. *Nat. Methods* 1–4 (2019).
19. Bastos, A.M. et al. Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
20. Schoenbaum, G., Roesch, M.R., Stalnaker, T.A. & Takahashi, Y.K. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat. Rev. Neurosci.* **10**, 885–892 (2009).
21. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
22. Sul, J. H., Kim, H., Huh, N., Lee, D. & Jung, M. W. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* **66**, 449–60 (2010).
23. Chudasama, Y. & Robbins, T.W. Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: Further evidence for the functional heterogeneity of the rodent frontal cortex. *J. Neurosci.* **23**, 8771–80 (2003).
24. Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N. & Komiyama, T. Area-specificity and plasticity of history-dependent value coding during learning. *Cell* (2019).
25. Saez, R.A., Saez, A., Paton, J.J., Lau, B. & Salzman, C.D. Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* **95**, 70–77 (2017).
26. Rikhye, R. V., Gilra, A. & Halassa, M.M. Thalamic regulation of switching

between cortical representations enables cognitive flexibity. *Nat. Neurosci.* **21**, 1753–1763 (2018).
27. Shuler, M.G. & Bear, M.F. Reward timing in the primary visual cortex. *Science (80-.).* **311**, 1606–1609 (2006).
28. Bari, A. et al. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology* **35**, 1290–301 (2010).
29. Isaacson, J.S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).
30. Neftci, E.O. & Averbeck, B.B. Reinforcement learning in artificial and biological systems. *Nat. Mach. Intell.* **1**, 133–143 (2019).

Figure Legends

Figure 1 | Lateral OFC-dependent reversal learning in a texture-discrimination task. a, *Top*: Schematic of experimental setup. *Bottom*: Trial-structure and outcome types (CR, correct rejection; FA, false alarm). **b**, Example of task performance during learning measured as mean correct rate (Hit + CR) and FA rate. After reaching stable high performance, stimulus-reward contingency was reversed ('rule-switch'). *Top*: Definition of salient task periods (*LN*: learning naïve, *LE*: learning expert, *RN*: reversal naïve, *RE*: reversal expert). **c**, Performance (*d'* values) in the four task periods pooled across 11 mice (different blue shadings). *Inset:* Number of sessions to reach expert level (*d'* > 1.5) for initial versus reversal learning. **d**, We expressed inhibitory DREADD (hM4Di) in S1 in 3 mice. Silencing S1 by systemic CNO application prevented learning (*d'* < 1.5 in *LE*; hence mice were not reversed). CNO-treated control mice (WT, n = 4) learned and re-learned normally. **e**, We expressed hM4Di in lOFC in 4 mice. Silencing lOFC during *RN* and *RE* impaired reversal learning. **f**, Silencing lOFC throughout all task phases did not affect initial learning but impaired reversal learning (n = 4 mice). OFC-silenced mice could still learn a new stimulus-outcome association (novel P600 go-texture). Data presented as mean ± S.E.M., *p < 0.05, **p < 0.01, ***p < 0.001, two-sided Wilcoxon rank-sum test. Box plots show median, 25th and 75th percentiles as box edges, and 5th and 95th percentiles as whiskers.

Figure 2 | *In vivo* Ca^{2+} imaging of lOFC and S1 neurons during reversal learning. a, *Top*: Schematic and photograph of cannula-window for imaging lOFC. *Bottom left*: Two-photon fluorescence image and GCaMP6f signals (ΔF/F) during different trial types for example lOFC L2/3 neurons imaged through a GRIN lens. *Bottom* right: Example Ca^{2+} transients during hit trials for an individual lOFC neuron with single-trial example of whisking-amplitude and lick events during a hit-trial below. B: baseline, S: stimulus-presentation window, R: reward-outcome window. **b**, Heat-map of single-trial ΔF/F responses of an example lOFC neuron (sorted by hit and CR; FA and misses not shown; performance (*d'*) indicated next to behavioural phases). **c**, Average Ca^{2+} transient amplitude in reward-outcome window for lOFC neurons for hit and CR trials (63 active out of 228 recorded neurons in 3 mice; n = 15 sessions). Across-trial average Ca^{2+} transients and percentage of active neurons for each phase shown above and below. **d**, *Top*: Schematic and photograph of cranial window above S1. We identified barrel cortex by whisker-evoked intrinsic imaging signals (two-photon imaging area indicated). *Middle and bottom left*: Fluorescence

image and GCaMP6f signals (ΔF/F) for example S1 L2/3 neurons. *Bottom* right: Example Ca^{2+} transients during hit trials for an individual S1 neuron, exhibiting responses during both stimulus-window and reward-outcome window, with single-trial example of whisking-amplitude and lick events below. **e**, Heat-map of ΔF/F transients for an example S1 neuron as in (b). **f**, Average Ca^{2+} transient amplitude in reward-outcome window for S1 neurons for hit and CR trials (261 active out of 539 recorded neurons in 5 mice; n = 56 sessions; 11 sessions discarded due to motion artefacts). S1 responses increased in hit trials of both expert phases (*LE* and *RE*). Across-trial average Ca^{2+} transients and percentage of active neurons for each phase shown above and below. Data presented as mean ± S.E.M.; *p < 0.05, **p < 0.01, two-sided Wilcoxon rank-sum test. Box plots show median, 25th and 75th percentiles as box edges, 5th and 95th percentiles as whiskers and dots as outliers.

Figure 3 | Distinct task-related dynamics of neuronal populations in lOFC and S1. a, Schematic illustrating 5 major classes of hit/CR selectivity changes upon rule-switch and their distribution in a 2D-scatter plot of selectivity before and after. To the right, dual assignment for *LE→RN* and *LE→→RE* comparison. We assessed selectivity by ROC analysis. **b**, Mean ΔF/F amplitude in the reward-outcome window for lOFC neurons for hit (left) and CR (right) trials, averaged across each salient phase. *Bottom*: Heat maps for 107 longitudinally imaged neurons (20 sessions in 3 mice). *Top*: Average values pooled across all neurons as box plots. **c**, 2D-scatter plot and marginal distributions (histograms) comparing hit/CR selectivity of lOFC neurons in **b** for *LE→RN* (*SI* computed in reward-outcome window). We display data points for neurons active only in *LE* above the plot, for neurons active in *RN* but not *LE* to the right. Active neurons with non-significant selectivity ($p > 0.05$, permutation test) are marked yellow. Note the high fraction of outcome-selective lOFC neurons. Neurons inactive in both phases are not included in the plot (percentage of active neurons on the right). **d**, Same plot as c but for *LE→→RE*. A fraction of lOFC outcome-selective neurons maintained their hit preference while another subset of previously inactive neurons acquired selectivity for the new-hit (51 active out of 68 chronically recorded neurons*;* 16 sessions in 3 mice). **e**, Same plot as in **b** but for S1 neurons (218 longitudinally imaged neurons; 28 sessions in 4 mice). **f**, Same *LE→RN* plot as in c but for S1 neurons. Most neurons retained their preference for the previous contingency (90 active out of 142 chronically recorded neurons; 20 sessions in 4 mice). **g**, Same plot as in f but for *LE→→RE.* A subset of neurons updated their outcome-selective preference in *RE* while another subset of previously inactive neurons acquired new selectivity for the newly rewarded hit trials (198 active out of

218 chronically recorded neurons; 28 sessions in 3 mice). Box plots show median, 25^{th} and 75^{th} percentiles as box edges, 5^{th} and 95^{th} percentiles as whiskers and crosses as outliers.

Figure 4 | Lateral OFC input reconfigures functional responses of S1 neurons. **a**, Retrograde AAV-retro/2-tdTomato injection, CLARITY and whole-brain imaging revealed long-range lOFC→S1 long-range projections (n = 2 mice; inset shows L2/3 lOFC). **b,** Left: Schematic of chronic imaging of S1 neurons in lOFC-silenced mice (*RN* and *RE*). Middle and right: 2D-scatter plots of *SI* values computed for *LE→RN* and *LE→→RE* together with marginal distributions as histograms (85 active neurons out of 164 neurons recorded in *LE* and *RN*, 24 sessions, one session discarded due to motion artefact; 115 neurons out of 210 neurons recorded in *LE* and *RE*, 25 sessions in 3 mice). **c**, Comparison of *SI* marginal distributions for *LE*, *RN*, and *RE* periods for lOFC neurons (Fig. 3 c,d), S1 neurons (Fig. 3 f,g), and S1 neurons in OFC-silenced mice (this figure, panel **b**). **d**, Heat-map of single-trial ΔF/F responses of an example lOFC neuron during *RN* sorted by hit and FA trials. Solid bars indicate periods for texture-presentation (light blue), reward (grey), and white-noise (red). **e**, Average Ca^{2+} transients (top) and mean ∆F/F amplitudes (bottom) of FA trials for lOFC neurons during four behavioural periods (63 active out of 228 neurons in 3 mice). *Inset*, Percentage of active neurons for hit and FA trials with overlap indicated. **f**, Average hit ∆F/F responses of two example outcome-selective neurons in S1 exhibiting trial-history dependent modulation with previous trial being rewarded (hit→hit; light grey trace grey) or punished (FA→hit; dark trace). **g**, Reward-history modulation index (RHMI) for outcome-selective neurons (blue) and neurons with acquired-selectivity (red) in lOFC, S1, and S1 in lOFC-silenced mice before (*LE*) and after (*RN*, *RE*) rule-switch. Data presented as mean ± S.E.M (*p < 0.05; bootstrap-permutation test; S.E.M. of RHMI with permutated indices as grey boxes). **h**, Schematic showing cortico-cortical feedforward (FF) and feedback (FB) interactions for value-prediction error computation in lOFC.

Methods

Animals. All experimental procedures were carried out in accordance with the guidelines of the Federal Veterinary Office of Switzerland and were approved by the Cantonal Veterinary Office in Zurich under license numbers 285/2014 and 234/2018. A total of 30 adult male mice (6-8-week old) were used in this study. For behavioural experiments, we used wild-type (WT) C57BL6/J mice (n = 16 mice). For imaging neurons in lOFC and S1, we used Rasgrf2-2A-dCre: CamK2a-tTA: TITL-GCaMP6f triple transgenic mice, expressing GCaMP6f in excitatory neocortical layer 2/3 neurons (n = 14 mice). For causal pharmacogenetic manipulations, both WT and L2/3-GCaMP6f animals were used (n = 3 WT mice and n = 3 GCaMP6f mice). To generate triple transgenic animals amenable to two-photon imaging, double transgenic mice carrying CamK2a-tTA (JAX# 016198[31]) and TITL-GCaMP6f (JAX# 024103[32]) were crossed with a Rasgrf2-2A-dCre line (JAX# 022864[33]). The de-stabilised Cre-recombinase expressed under the control of the Rasgrf2-2A promoter was stabilised by trimethoprim (TMP, Sigma T7883) to render it functional. TMP was reconstituted in Dimethyl Sulfoxide (DMSO, Sigma 34869, 100 mg/ml), freshly prepared before each induction, and administered two weeks before surgery. During induction, mice were given a single intraperitoneal injection (150 mg TMP/g body weight diluted in 0.9% saline solution) using a 29g needle. To specifically label and image from S1→lOFC projection neurons, we injected AAV2.9.hSyn.FLEX.GCaMP6f virus into S1 of WT mice. Mice were grouped with their WT siblings and housed at 24°C and variable humidity in 12-hour reverse dark-light cycle (7:00 a.m. to 7:00 p.m.). At the end of an experiment, the animals were deeply anesthetised and transcardially perfused or euthanised with an overdose of pentobarbital (150 mg/kg body weight, i.p.). All efforts were made to minimise suffering. All mice belonged to the C57BL6/J strain.

Reversal learning task. Mice were extensively handled during pre-training sessions to familiarize them with the experimenter and experimental setup. Once they had acclimatised to handling, mice were placed on water-restriction and trained on a go/no-go tactile-discrimination task. Mice remained on water-restriction for the remainder of the experiment. The behaviour set-up has been described previously[10]. During the start of each trial, an auditory cue (2 beeps at 2 kHz, 100 ms duration with 50 ms interval), indicated the approach of one of two possible textures (sandpapers of grit size P100, rough texture; P1200, smooth texture). The texture was positioned to reach the mouse's whiskers and 'go' or 'no-go' textures were presented pseudo-

randomly with no more than three consecutive repetitions. The texture stayed in touch with the whiskers for one second ('sensation'), after which it moved out of reach. An additional auditory tone (response cue; 4 beeps at 4 kHz, 50-ms duration with a 25-ms interval) signalled the start of a 2-second 'response window' during which the mouse had to lick or withhold from licking the water sprout to indicate its choice ('outcome or response', 2 seconds). A sucrose-water reward was delivered only for licks in response to the 'go' texture and after the response cue ('hit'). Incorrect licks in response to the non-target 'no-go' texture ('false alarms', FA) was punished with a brief period of mild auditory white noise. Reward and punishment were omitted when mice withheld licking for the no-go ('correct-rejections', CR) or go ('miss') textures. The licking detector remained in a fixed and reachable position throughout the entire trial. Animals were motivated to perform the task and typically showed a fraction of 10-15% miss trials during *LN* period which reduced significantly upon learning (*LE*) and remained same upon rule-switch.

Mice proficiently performed the sensory-discrimination task from learning naïve (*LN*) through expert phase (*LE*). Once mice had achieved stable performance of the tactile-discrimination task (reaching a *d' = 1.5* for 3-4 sessions), the stimulus-response mapping was switched ('rule-switch'). Upon rule-switch, performance initially dropped to chance level or below. However, after 4-5 days, all mice (n = 11 out of 11 mice) learned the new texture-response mapping and increase performance from reversal naïve (*RN*) through expert phase (*RE*) as quantified by the increase in the discriminability index (*d'*) (training period 4-5 days, 200-300 trials/session/day).

Animal training and performance measurement. We quantified mice task performance using the discriminability index d-prime (*d'*) rather than percent correct to account for motivation and criterion[34]. We set the learning threshold to *d'* = 1.5. d' was calculated for each session as = Z(hit/(hit+miss)) − Z(FA/(FA+CR)) with $Z(p)$, $p \in [0,1]$, being the inverse of the cumulative Gaussian distribution (FA, number of false alarm trials; CR, number of correct rejection trials). We selected in both training periods pre- and post-reversal two relevant phases corresponding to the salient phases - learning and reversal naïve (*LN* and *RN*, respectively), in which the mice were performing lower or close to chance level ($d' = 0$, $p < 0.05$ for $d' > 0$, n = 1-3 sessions), and learning and reversal expert (*LE* and *RE*, respectively, n = 1-3 sessions), in which the mice were stably performing above a criterion set as $d' = 1.5$. Expert sessions were always selected from the last sessions available immediately before rule-switch (*LE*) or task completion (*RE*), and this resulted in high

performance level (d' > 2). For imaging data, only days among these respective sessions were used.

Whisking and licking measurement. During task performance, whisker kinematics and fine body movement were simultaneously monitored using high-speed cameras. We identified behavioural correlates of task learning by quantifying licking rate and whisking amplitude obtained from lick-sensor measurements and high-speed videography, respectively. The whiskers were illuminated with 940 nm infrared LED light and movies were acquired during the behaviour at 500 Hz (500 × 500 pixels) using a high-speed CMOS camera (A504k; Basler). Average whisker angle across all imaged whiskers was measured using automated whisker tracking software. The whisking amplitude (envelope) was calculated as the difference in maximum and minimum whisker angle along a sliding window equal to the imaging frame duration (142 ms). Principal whisker velocity was calculated by applying a band-pass filter to the whisking angle time vector and then computing its first derivative. For all trials recorded (n = 3 mice), the first and last possible time point for whisker-to-texture contact was quantified manually through visual inspection.

Licking was detected by using a piezo-electric sensor attached to the lick spout and lick rates were calculated by thresholding this signal and counting the number of events per unit of time. Multiple consecutive threshold crossings which occur in rapid succession can result in a lick rate that exceeds the physical capability of a mouse. We therefore made the reasonable assumption of a peak lick rate of 10 Hz based on manual checks on videography. A low pass filter was applied to the lick rate time series, which effectively combined multiple events occurring within a 100 ms window into one event. Expert mice showed a decrease of early licks. While early licks are not exhibited immediately upon rule-switch when the behavioural performance is low, lick rates are slightly lower compared to expert sessions.

Open-field test. General locomotor activity was measured in an open-field (a rectangular arena of 40 x 30 x 20 cm)[35] made from grey Plexiglas that was illuminated from a centred diffuse light source. A single animal was exposed to the environment for 5 minutes while being recorded by a video camera placed above the open field and operated by LabVIEW (National Instruments). Mouse velocity (cm/s) and distance covered (cm) were analysed using the EthoVision software.

Horizontal ladder-rung test. A 1-m long horizontal ladder, consisting of two platforms connected by an irregular pattern of 70 rungs was used. The distance

between rungs varied between 1-3 cm. Mice were given time to practice with three trials before being tested. Three trial sessions per animal were recorded using a high-speed camera (Nikon AF Nikkor) at 100 frames per second. Each forepaw placement was analysed and the quality of the placement was scored using the following scoring system[36]. A perfect paw placement on the rung was scored as 1; partial digit placement, correction and replacement were scored as 0.5, slip or total miss were scored as 0. The success rate was calculated for each animal group as

$$\textit{Success rate} = (\textit{Total score}/\textit{Number of steps}) \times 100 \quad (1)$$

Virus injection. Mice were briefly anaesthetised with isoflurane (2%) in oxygen in an anaesthesia chamber and subsequently transferred to a stereotactic frame (Kopf Instruments). Body temperature was maintained at ~37°C using a heating blanket with a rectal thermal probe. The eyes of the mouse were covered by Vitamin A cream (Bausch & Lomb) during the surgery. The cranium was secured with ear bars and anaesthesia was maintained during the surgery with 0.8-1.2% isoflurane. After disinfection with Betadine, the skin was opened using a scalpel and an L-shaped incision was made in the skin, and the cranial surface was cleaned using absorbent swabs (Sugi; Kettenbach GmbH). We identified lOFC based on stereotactic coordinates from previous studies (2.6 mm anterior and 1.2 mm lateral from bregma)[13]. For S1, injection coordinates were 3.5 mm lateral and 1.5 mm posterior from bregma. The skull was thinned along a 1-mm line at the rostral edge of S1 using a Dremel drill with occasional cooling with saline. After drilling through the cranium, the dura was punctured using a glass micropipette filled with the virus suspended in mineral oil. Several injections (3-4) were made at neighbouring sites, at a depth of 200-250 µm. A volume of 100-150 nl of virus was injected at 50 nl/min rate at each site. After each injection, the pipette was held in place for 5-8 minutes before retraction to prevent leakage. Skin was sutured using a synthetic, monofilament, non-absorbable suture (Prolene 7.0, Ethicon).

Cranial window and GRIN lens implantation. To study neural dynamics in the lOFC, a chronically implanted metallic cannula was implanted on top of lOFC with a glass coverslip at its base. Cannula implantation and cranial window preparation was performed under isoflurane anaesthesia following details as described above. A circular piece of cranial bone (diameter ~ 1.5 mm) was drilled on top of OFC using a Dremel drill. A modified biopsy punch (diameter 1.0 mm; Miltex) was inserted 1.5 mm deep into the cortical tissue for two minutes. The cortical tissue (primary and

589 secondary motor areas) was gently aspirated with a cut using a 27-gauge needle
590 connected to a water jet pump, while constantly being rinsed with Ringer solution.
591 We removed the overlying cortex using aspiration until layer 5 (depth 1.5-1.7 mm)
592 and implanted a stainless-steel cannula (internal diameter 1.0 mm, 1.5 mm height)
593 was with its base covered by a cover glass (0.17 mm thickness) 1.6-1.8 mm below
594 the pial surface. The cannula was secured in place by UV curable dental acrylic
595 cement (Ivoclar Vivadent). We waited two-three weeks after surgery before
596 commencing training. Before each imaging session, a rod-like gradient-index (GRIN)
597 lens (NEM-100-48-00-50-NC, customised needle endomicroscope for two-photon
598 microscopy, ~ 0.4 pitch, corrected for wavelength λ = 920 nm, diameter = 1.0 mm,
599 length ~ 4.3 mm; GRINTECH GmbH, Jena) was inserted through the cannula and
600 neurons were imaged 100-300 um below. Before each imaging session, the cannula
601 was cleaned with distilled water.
602 To allow long-term *in vivo* calcium imaging in S1, a cranial window was
603 implanted over S1 as described previously[10,37]. A metallic head-post for head fixation
604 was glued to the skull, contralateral to the cranial window, using dental acrylic. One
605 week after chronic window implantation, mice were handled daily for one week while
606 they became acclimatised to a minimum of 15 mins of head-fixation.
607
608 **Brain clearing and light-sheet microscopy.** To verify task-relevant projections and
609 connectivity between S1 and lOFC, we injected retrograde AAV-retro/2-shortCAG-
610 tdTomato virus *in vivo*. Two to three weeks after virus injection, animals were
611 perfused, and the brains entered a clearing protocol using CLARITY[38]. After
612 perfusion, the brains were post-fixed for 48 hours in a hydrogel solution (1%
613 paraformaldehyde, 4% acrylamide, 0.05% bis-acrylamide, 0.25% VA044)[38,39] before
614 the hydrogel polymerization was induced at 37°C. Following the polymerization, the
615 brains were immersed in 40 ml of 8% SDS and kept shaking at room temperature
616 (RT) until the tissue was cleared sufficiently (20-40 days depending on the age of the
617 animals). Finally, after 2-4 washes in PBS, the brains were put into a refractive index
618 matching solution (RIMS)[38] for the last clearing step. They were left to equilibrate in 5
619 ml of RIMS for at least 4 days at RT before being imaged.
620 Cleared brains were imaged using a mesoSPIM light-sheet microscope
621 (www.mesospim.org)[18]. Whole-brain imaging revealed that lOFC receives direct
622 monosynaptic bottom-up, feed-forward projections from both superficial (L2/3) and
623 mostly deep (L5 and L6) layers of S1. Conversely, a similar injection in mouse S1
624 (2.55 mm posterior and 3.5 mm lateral from bregma)[10] revealed superficial cortical
625 L2/3 neurons in mouse S1 receiving direct top-down feedback projections from lOFC.

CNO application. Inhibitory DREADDs (CaMKIIα-hM4D(G_i)-mCherry) were used in the chemogenetic silencing experiments and neuronal populations of interest were virally transfected with AAV-hM4Di injected unilaterally on the superficial layers (L2/3) of contralateral lOFC and bilaterally to superficial (L2/3) and deeper (L5) layers of S1. Intraperitoneal (i.p.) injection of clozapine *N*-oxide (CNO-dihydrochloride, 1-5 mg/kg, TOCRIS, Cat.No.4936), the ligand that activates hM4Di, silenced the activity of neurons. Clozapine (1-5 mg/kg) was used as control as there are reports that a small proportion of systemically-administered CNO is metabolized to clozapine[40].

***In vivo* electrophysiological recordings.** We characterised pharmacogenetic silencing of lOFC neurons by performing acute, *in vivo* electrophysiology in a subset of hM4Di-injected animals after completion of the reversal learning protocol. To perform acute recordings, animals were anesthetised with isoflurane (2% for induction and 0.8% during recording), and their body temperature was maintained stably using a heating pad. A small craniotomy (1-mm diameter) was performed to provide access to the left OFC and the brain was covered with silicon oil. A silver wire was placed in contact with the CSF through a small trepanation (0.5 mm) over the cerebellum to serve as reference electrode. A silicon probe (Atlas Neurotechnologies, 16 linear sites, 100 µm spacing) was implanted through the craniotomy into the left cortical hemisphere and we recorded multi-unit activity from the injection site in the left OFC and surrounding cortex. We waited 30 minutes to allow the recording to stabilise after implantation of the electrode array. After stabilisation, the broadband voltage was amplified and digitally sampled at a rate of 30 kHz using a commercial extracellular recording system (RHD2000, Intan Technologies). The raw voltage traces were filtered offline to separate the multi-unit activity (MUA; bandpass filter 0.46-6 kHz) using a fourth-order Butterworth filter. Subsequently, the high-pass data were thresholded at 6.5 times the standard deviation across the recording session and the numbers of spikes in windows of interest were counted. After a baseline recording of 30 mins, CNO (1-5 mg/kg) was injected (i.p.). During the baseline period (30 minutes), the average firing rate remained stable, while upon CNO injection the average firing rate in the lOFC steadily decreased over time. Recording electrodes in the lOFC showed a stable and significant decrease in spiking activity 30 minutes after CNO administration, while control electrodes from areas uninfected by the virus did not show any modulation. To combine data across mice, the activity at sites with clear MUA was expressed in percent of the baseline value, i.e. the average spike rate during the 30-minute pre-

injection baseline (100%). All multi-units were then combined from the injected or control region and a t-test was performed between the baseline period (-30-0 minutes pre-injection) and the post injection period (30-60 minutes post injection).

Intrinsic signal optical imaging. The S1 barrel cortex was identified using intrinsic signal optical imaging under approximately 0.8-1 % isoflurane anaesthesia. The cortical surface was illuminated with a 630-nm LED, multiple whiskers were stimulated (2 to 4 rostro-caudal deflections at 10 Hz), and reflectance images were collected through an objective with a CCD camera (Toshiba TELI CS3960DCL; 12-bit; 3-pixel binning, 4273 347 binned pixels, 8.6-mm pixel size, 10-Hz frame rate)[41].

Intrinsic signal changes were computed as fractional changes in reflectance relative to the pre-stimulus average (50 frames; expressed as DR/R). The centres of the barrel columns corresponding to stimulated whiskers were located by averaging intrinsic signals (15 trials), median-filtering (5-pixel radius) and thresholding to find signal minima. Reference surface vasculature images were obtained using 546-nm LED and matched to images acquired during two-photon imaging.

Two-photon imaging. We used a custom-built two-photon microscope controlled by HelioScan[42], equipped with a Ti:Sapphire laser system (approximately 100-femtosecond (fs) laser pulses; Mai Tai HP; Newport Spectra Physics), a water-immersion 16X Olympus objective (340LUMPlanFI/IR, 0.8 numerical aperture, NA) for S1 imaging and a 20X Leica objective (Leica Plan Apo 0.6 NA) for GRIN lens based OFC imaging, galvanometric scan mirrors (model 6210; Cambridge Technology), and a Pockels Cell (Conoptics) for laser intensity modulation.

Based on intrinsic imaging, along with the blood vessel pattern, we targeted specific areas of interest for two-photon imaging of L2/3 neurons in each mouse. We excited GCaMP6f at 940 nm and detected green fluorescence with a photomultiplier tube (Hamamatsu). Images (128x64 pixels) were acquired at 12-Hz frame rate and 10-50 cells per field of view were imaged simultaneously. Single trials of 6-8 s duration were recorded, with 1-s breaks between trials to allow the data to be written to hard-disk during inter-trial periods.

Calcium imaging analysis. Calcium imaging data was first motion corrected using an online piecewise rigid 2d (planar) method (NoRMCorre: Non-Rigid Motion Correction) in MATLAB (Mathworks). Regions of interest (ROI) corresponding to individual neurons were found from both the mean image and the standard deviation image generated from a single-trial time series using ImageJ (US National Institutes

of Health). ROI masks were manually selected using an online method (OCIA) in MATLAB and raw fluorescence time courses (F(t)) were then extracted as the (non-weighted) mean pixel value for each ROI. Another fluorescence time course was extracted from a neuropil defined by an ROI selecting a portion of non-somatic tissue in the imaging frame. The neuropil calcium signal never resulted in activity peaks significantly high to be classified as an active neuron (check Criteria for active neurons). The background was subtracted on each channel (bottom first percentile fluorescence signal across entire time series). A running estimate of fractional change in fluorescence time courses was calculated by subtracting the baseline fluorescence $F_0(t)$ from F(t), then dividing by $F_0(t)$

$$\Delta F/F(t) = (F(t) - F_0(t))/F_0(t) \quad (2)$$

$F_0(t)$ was estimated as the mean fluorescence value of the first 1.5 s prior to tactile stimulus onset. For cells that were not silent in the pre-stimulus window, $F_0(t)$ was instead taken as 8^{th} percentile of a trailing 1.5-s sliding window.

Alignment of cell-masks across days. All analyses for the alignment of cell-masks across days were manually performed with the aid of custom MATLAB GUIs in the OCIA software. To align masks across any pair of daily sessions, we first chose one set for the first day and then imported it onto the single-trial image series of the subsequent days. When displacement occurred, the masks were manually moved to the corresponding neurons. This was done for all pairwise combinations of days. We then manually observed by eye each ROI mask confronting it to both the mean and the standard deviation image of the time series on ImageJ, to confirm the presence of each cell across days. If the z-plane did not match and a cell was not found, it was excluded from further longitudinal analysis.

Criteria for active neurons. To determine if a neuron was active during a time-period of interest (stimulus-related and reward-outcome related responses), we independently tested its evoked response using conservative criteria. For each neuron, we calculated its mean response and its peak value ($\Delta F/F$) during the 0.9 s window after a texture was presented (i.e. for stimulus presentation-window) or during the1.6-s window after the texture was removed (i.e. for reward-outcome window). A neuron was considered active if all the following criteria were met:

- its response was significantly ($p < 0.01$, *t*-test) different from the average pre-stimulus baseline response (1.5 s before texture is presented).

- its mean response (for stimulus-presentation or reward-outcome window) was more than 3*noise from the baseline. This baseline was calculated by averaging a 35-point sliding-window across the trial response and taking the 5th percentile of the mean response distribution. The noise level taken as the 1st percentile of the distribution of the standard deviation calculated across the same sliding window.
- its peak response (ΔF/F) (for stimulus or reward-outcome window) was greater than 25%.
- In the 2D scatter plots of selectivity indices (see below) neurons were considered active if they were active in either of the considered learning periods (e.g. *LE* and *RN*). In other words, they were considered inactive only if they were inactive in both respective periods.

Selectivity index. We assessed the selectivity of single-neuron activity for specific trial-types using a receiver operating characteristic (ROC) analysis, which quantifies the ability of an ideal observer to discriminate between trial-types based on single-trial responses[16,10]. For the purpose of this study, we assessed selectivity for hit vs. CR trials. We performed the ROC analysis on the segments of the ∆F/F transients in the trial period of interest, i.e., either in the 2-s long reward-outcome window or in the 1 s long stimulus window. Specifically, each trial was assigned a ‘‘discrimination variable’’ score (*DV*) equal to the dot product similarity of the ∆F/F segment to the mean ∆F/F segment for the same trial-type minus the dot-product similarity to the mean for the other trial-type (see also **Extended Data Fig. 8**). Thus, we computed for hit trials

$$\boldsymbol{DV_{Hit}} = \boldsymbol{H_i}(\overline{\boldsymbol{C}}_{\forall j \neq i} - \overline{\boldsymbol{C}}) \quad (3)$$

and for CR trials

$$\boldsymbol{DV_{CR,i}} = \boldsymbol{C_i}(\overline{\boldsymbol{H}} - \overline{\boldsymbol{C}}_{\forall j \neq i}) \quad (4)$$

where H_i and C_i are the single-trial ∆F/F segments for the *i*-th hit and CR trial, respectively, and $\overline{\boldsymbol{H}}$ and $\overline{\boldsymbol{C}}$ denote the mean ∆F/F segments for the respective trial type (excluding the individual trial under consideration). We classified trials as belonging to the go-texture or the no-go-texture if *DV* (DV_{Hit} or DV_{CR}) was greater than a given criterion. To determine the fraction of trials an ideal observer could correctly classify, we constructed an ROC curve by varying this criterion value across the range of *DV*. At each criterion value, we plotted the probability that a hit trial exceeded the criterion value against the probability that a CR trial exceeded the

criterion value. The area under this ROC curve (AUC) indicates the selectivity for trial type, with an AUC value of 0.5, meaning no selectivity. We defined the "selectivity index", *SI*, such that it spanned the range from -1 (CR-preferring neurons) to +1 (hit-preferring neurons) by calculating

$$\boldsymbol{SI = 2 \times (AUC - 0.5)} \quad (5)$$

We tested whether neurons showed trial-type selectivity above chance using a permutation test creating 500 permutations with trial-type labels randomly shuffled. From these permutations, we created a distribution of indices that could have arisen by chance and considered a neuron's *SI* value as significant if it fell outside the centre 95% interval of this distribution ($p < 0.05$).

Functional classification of neurons. Neurons that met the activity criterion in at least one of the salient learning periods were classified in different groups according to their hit/CR *SI* value changes upon rule-switch. For each of these neurons we compared the *SI* value in the pre-reversal period (*LE*) to the *SI* value in the two post-reversal periods (*RN* and *RE*). This resulted in two classifications for each neuron (for *LE*→*RN* comparison and *LE*→→*RE* comparison) (**Fig. 3a**). When two *SI* values before and after reversal were found concordant, i.e. of the same sign and significant, a neuron's response was classified as 'outcome-selective' for the respective post-reversal phase and the specific trial time-window considered (stimulus or reward-outcome). Such neuron's response amplitude was significantly higher for hit compared to CR trials (or CR compared to hit trials) independent of stimulus-identity (in the 2D scatter plots these neurons are found in the upper right and lower left quadrants). When *SI* values before and after reversal were discordant, i.e., of opposite sign and significant, the neuron's response was classified as 'stimulus-selective' as it switched from hit- to CR-preferring (or CR- to hit-preferring), where the new CR was associated with the same stimulus as the previous hit. In the 2D scatter plot these neurons are found in the upper left and lower right quadrants. If an active neuron was discriminating above chance during the pre-reversal period *LE* and lost significant selectivity in the pre-reversal period considered (*RN* or *RE*), or if it simply became inactive, it was classified as a 'lost-selectivity' neuron. Likewise, if an inactive neuron or an active neuron without significant selectivity in the pre-reversal period became active and gained a significant selectivity for the new hit/CR trials, it was included in the 'acquired-selectivity' group. Finally, all the active neurons that did not show a significant *SI* value during either phase (based on permutation tests), were considered 'non-selective'. Each one of these neurons was assigned twice to a

functional group, in earlier (*RN*) and later phases of reversal (*RE*). We tracked the class transition through the course of re-learning using a fate map. For each *LE*→*RN* group we showed the fraction of neurons falling into the new *LE*→→*RE* classes. Only active neurons during both phases are shown.

Reward-history modulation index. To quantify the effect of previous performance on neural responses, we analysed how response magnitude varied as a result of the outcome of the previous trial (punishment or reward)[17]. We compared the response magnitude of each neuron during a hit trial when the previous trial was rewarded hit ($R_{Hit\text{-}Hit}$) versus the response magnitude when the previous trial was punished ($R_{FA\text{-}Hit}$). To quantify modulation by previous trial history, we created a reward-history modulation index (RHMI) by normalizing the difference between these two history-dependent responses by the mean overall response of all the Hit trials:

$$RHMI = \frac{|R_{FA-Hit} - R_{Hit-Hit}|}{R_{Hit}} \quad (6)$$

Only cells that were active during a specific phase were included in the RHMI analysis for that respective phase. To check whether a neuron was modulated above chance, a bootstrap permutation test was performed (500 permutations).

Generalized linear model. To estimate the contribution of behavioural and task variables (cue, stimulus onset and offset separated by behavioural response, reward delivery, punishment, licking) to the activity of each neuron, we fit a Poisson generalized linear model (GLM) for each session (MATLAB glmnet package). We first down-sampled deconvolved neural data and all behavioural and task variables to 10 Hz and then smoothed neural activity using a Gaussian filter. Regression functions were created from behavioural and task variables by implementing vectors of Gaussian filters (all filters had a standard deviation of 1 s, overlapping and evenly distributed, 1 Gaussian/3 frames, 100 ms/frame, 144 filters). Each imaging session consisted of 100-120 trials of 6 seconds each (15 Hz) (training set 75% of each run, testing sets 25%; 10-fold cross validated with 11 evenly spaced chunks of trials). We used an elastic net regularization consisting of 99% L2 and 1% L1 methods for each individual neuron. Deviance explained was calculated by comparing the activity predicted by the model to the actual activity calculated using data not used during the fitting procedure. Finally, the contribution of each variable to the neural activity is derived by calculating again the deviance explained using just that variable and

normalizing it to the total deviance explained. This is plotted separately for each group of neurons.

Statistical analysis. Statistical analyses are described in the main text and in figure legends. If not stated otherwise, we used non-parametric statistical analyses (two-sided Wilcoxon rank-sum test) or permutation tests to avoid assumptions about the distributions of the data. When assumptions could be made based on previous literature and on small datasets (Fig. 1d, Extended Data Fig. 1c and 5), *t*-test was used. All statistical analysis was performed using custom written routines in MATLAB. Quantitative approaches were not used to determine if the data met the assumptions of the parametric tests.

Extended Data Figure Legends

Extended Data Figure 1 | S1-dependent tactile-discrimination-based reversal learning task. a, Time-course of task-performance (discriminability index, *d'*) of individual mouse reveals dynamics of learning and reversal learning upon rule-switch. Each line in various blue shades represents a single mouse of a total of 11 mice. **b,** Percentage of correct decision '(hit+CR)/all trials' as 'outcome rate' plotted during the four salient behavioural phases of learning (learning naïve, *LN*; learning expert, *LE*) and reversal (reversal naïve, *RN*; reversal expert, *RE*) (n = 11 mice). **c,** Reversal performance is stable and remains high when mice with reversed reward contingency (P1200 as go-texture, *RE*) were tested 6 weeks later (n = 2 mice). **d,** Reversal learning is independent of initial texture training (fine grit size sandpaper P1200 texture as initial go-texture; n = 2 mice). **e,** Texture-discrimination is dependent on sensory input. *Left:* Keeping textures out of reach in expert mice after reversal (*RE*) impaired their performances (n = 3 sessions in two mice). *Right:* Clipping whiskers in expert mice similarly resulted in impaired performance (low *d'*) indicating sensory input is essential for the correct execution of the task (n = 3 mice, longitudinally studied before and after whisker-clipping). Data presented as mean ± S.E.M., *p < 0.05, **p < 0.01, ***p < 0.001, two-sided Wilcoxon rank-sum test.

Extended Data Figure 2 | Whisking and licking behaviour during reversal learning. a, *Upper row*: Time-course of envelope whisking amplitude aligned to first-touch during go- (left) and no-go-trials (right) across two salient periods of initial learning (learning naïve, *LN*; learning expert, *LE*). In naïve animals (*LN*), mice exhibited low amplitude whisking activity throughout most of the trial. In expert mice (*LE*), whisking behaviour became time-locked to the arrival of the texture. *Lower row*: equivalent whisking traces for the periods after rule-switch (reversal naïve, *RN*; reversal expert, *RE*; *right*). Both in *RN* and *RE* periods, mice showed stimulus time-locked whisking amplitude (n = 3 mice). Note that amplitudes and temporal profiles of the whisking envelope were similar for the smooth P1200 and the rough P100 texture, independent of stimulus-outcome association. **b,** Equivalent analysis as in (a) but for the mean whisking velocity. No significant difference was found in the velocity profile between the two textures in the stimulus-presentation window. **c,** Time-course of average lick rates during go-trials across two salient phases of initial learning (left) and reversal learning (right) (n = 11 mice). Expert mice (*LE* and *RE*) showed both an increase in licking activity during report window (grey) and a

decrease of early licks (B-baseline, S-stimulus-presentation, R-reward). Data is presented as mean (solid line) ± S.E.M. (shaded area).

Extended Data Figure 3 | Immunohistochemical and behavioural validation of pharmacogenetic silencing using hM4Di. a, Neuronal silencing was achieved via viral injection of inhibitory DREADD (AAV-hM4Di-mCherry) into S1 and/or lOFC in mice followed by systemic CNO application. S1 injection (top) was bilateral and lOFC (LO) injection (below) was unilateral and to the ipsilateral side of the barrel field. **b,** Injection of hM4Di in lOFC and systemic administration (i.p.) of clozapine (1-5 mg/kg) after rule-switch (*RN* and *RE*) selectively impaired reversal learning (n = 3 mice). **c,** Injection of hM4Di in lOFC and CNO treated animals showed increased perseverative errors (false alarm, FA) in *RE* compared to *LE* (n = 4 mice). **d-e,** Silencing medial OFC (MO) by injecting hM4Di unilaterally in the MO, followed by daily systemic CNO application after rule-switch (*RN* through *RE* period), did not have any effect on reversal learning. *p < 0.05, **p < 0.01, ***p < 0.001 two-sided Wilcoxon rank-sum test. Data is presented as mean ± S.E.M.

Extended Data Figure 4 | Electrophysiological validation of lOFC silencing using hM4Di. a, Timeline depicting experimental sequence for validation of lOFC (LO) silencing (top). Schematic of acute electrophysiological recording from frontal cortex (bottom). DAPI stained slice imaged with a confocal microscope showing red fluorescence from DiD to mark the probe location. Example traces from three electrode contacts from one recording session for pre- and post-CNO injection (middle). Box plots showing change in firing rate (% change relative to baseline) for electrode contacts above, in, or below lOFC. Plots show median, 25^{th} and 75^{th} percentiles as box edges, and 5^{th} and 95^{th} percentiles as whiskers. To the right, example waveforms from units showing significant modulation by CNO. *p < 0.05, *t*-test.

Extended Data Figure 5 | Unaltered whisking and simple behaviour following OFC cannula implantation. a, A schematic diagram and whole-brain image showing the location of cannula implantation in OFC. Coloured regions on the schematic indicate pre-motor and motor areas as described in the previous studies[43,44,45,41] (left hemisphere), or regions according to the Allen institute Common coordinate framework (right hemisphere). **b,** A schematic diagram based on the Allen brain atlas, light-microscopic and confocal view shows the GCaMP6f expressing mice in lOFC (LO) and cannula placement above the virus injection site. **c,** Whisking

behaviour in OFC cannula-implanted animals is preserved. Envelope whisking amplitude (top) and whisking velocity (bottom) in expert animals (*RE*) centred on the texture-approach (n = 2 mice). **d,** Open-field test showed normal locomotor function of wild-type and OFC cannula-implanted mice (n = 4 WT and n = 2 OFC cannula-implanted mice). Representative picture of locomotor track (top) and heat-map (bottom) of an OFC cannula-implanted mouse. Total distance covered (cm) and mean velocity (cm/s) is plotted. Scale bar = 5 cm. **e,** Horizontal ladder-rung test showed normal locomotor function of wild-type (WT, n = 4) and OFC cannula-implanted mice (n = 2). A representative picture showing paw placement of a mouse on irregular horizontal rung-ladder. **f,** Analysis of paw placement of the limb contralateral to the cannula-implanted side showed no significant difference between WT and OFC cannula-implanted mice. **g,** No differences were seen in paw placement of the limb ipsi- or contralateral to the cannula-implanted side in OFC cannula-implanted and in control WT mice. Data is presented as mean ± S.E.M.

Extended Data Figure 6 | Re-learning task with neutral context and *in vivo* Ca^{2+} imaging of lOFC neurons. a, Schematic of the stimulus-outcome associations in a three-textures task with positive (large reward), neutral (small reward), and negative (punishment) context. Same coarse P100 and smooth P1200 sandpapers were used, but an additional intermediate coarseness P600 sandpaper was introduced as go-neutral context (go_{nc}) associated with a small reward, that did not change upon reversal. **b**, Average Ca^{2+} transient amplitude in the reward-outcome window for lOFC neurons for Hit, Hit_{nc} and CR trials (n = 63 active neurons out of 228 neurons recorded in three mice; n = 15 sessions) showing increased Hit responses upon rule-switch but no significant changes during Hit_{nc} trials. Across-trial average Ca^{2+} transients for each behavioural period are shown above. All box plots show median, 25^{th} and 75^{th} percentiles as box edges, and 5^{th} and 95^{th} percentiles as whiskers.

Extended Data Figure 7 | Task-related functional dynamics in S1→lOFC projecting neurons during reversal learning. a, Retrograde AAV-retro/2-tdTomato injections *in vivo* in the lOFC followed by clearing the brain using CLARITY and whole-brain light-sheet microscopy revealed feed-forward S1→OFC projections from both deeper (L5 and 6) and superficial (L2/3) layers of S1 (n = 2 mice). Labelling is weaker on the contralateral side of the injection site. **b,** S1→lOFC projecting neurons were labelled with GCaMP6f using a dual-viral strategy with retrograde AAV2-retro/2-Cre injected in lOFC and Cre-dependent AAV-DIO-GCaMP6f in S1. *Inset*, L2/3 neurons in S1 labelled with such strategy. **c,** Average Ca^{2+} transient amplitude in the

reward-outcome window shows a significant increase in response amplitude during expert phases of training (*LE* and *RE*) (n = 96 active neurons over n = 135 recorded neurons in two mice, n = 5 sessions/phase). **d,** *Top,* S1→lOFC projecting neurons were labelled using a dual-viral strategy with retrograde AAV2-retro/2-Cre injected in lOFC and Cre-dependent AAV-DIO-GCaMP6f in S1. *Bottom*, peak responses of S1→lOFC projection neurons averaged across hit (left) and CR (right) trials, longitudinally measured across four salient periods (n = 96 neurons from n = 2 mice, n = 5 sessions/phase). Box plots (median, red line; 25th and 75th percentile, box edges; whiskers as most extreme non-outliers; outliers, red crosses; zero, dashed grey line) are also shown (*inset*). **e,** Scatter plot and histogram comparing selectivity index (*SI*) of S1→lOFC projecting neurons during learning expert (*LE*) and reversal naïve (*RN*) phase (n = 39 active neurons over n = 46 neurons from n = 2 mice, n = 5 sessions/phase). **f,** Scatter plot and histogram comparing *SI* of S1→lOFC projecting neurons during *LE* and reversal expert (*RE*) phase (n = 61 active neurons over n = 73 from n = 2 mice, n = 5 sessions/phase). All box plots show median, 25th and 75th percentiles as box edges, and 5th and 95th percentiles as whiskers. Data presented as mean ± S.E.M., *p < 0.05, **p < 0.01 two-sided Wilcoxon rank-sum test.

Extended Data Figure 8 | Tracking neuronal responses during early and late phases of reversal learning. a, A schematic view of the step-by-step derivation of the selectivity index (*SI*) from the ROC curves. **b**, Selectivity indices of longitudinally tracked lOFC neurons across the salient task-periods of *LE*, *RN*, and *RE*. Marker colours for *RN* and *RE* indicate the assigned classes for the *LE→RN* and *LE→→RE* comparisons, respectively. Plots are shown separately for each *LE→RN* class. **c**, Fate mapping of longitudinally tracked lOFC neurons. For each *LE→RN* assigned class, the distribution of these neurons across classes for the *LE→→RE* comparison is shown as coloured bar on the right. **d**, Same as in (b) but for S1 neurons. **e**, Same as in (c) but for S1 neurons. **f**, Same as in (b) but for S1 neurons in lOFC-silenced mice. **g**, Same as in (c) but for S1 neurons in lOFC-silenced mice. *Inset* in **e**, The fate distributions of the non-selective neurons in *LE→RN* show a significantly smaller fraction of neurons that acquire selectivity for the newly rewarded go-texture in the *RE* phase in S1 neurons when lOFC was silenced in mice (22% vs. 60%, one-tailed Chi-square test). Note that the fate mapping plots include additional neurons compared to (b), (d), and (f) as these were not assigned an *SI* value in each phase but still classified.

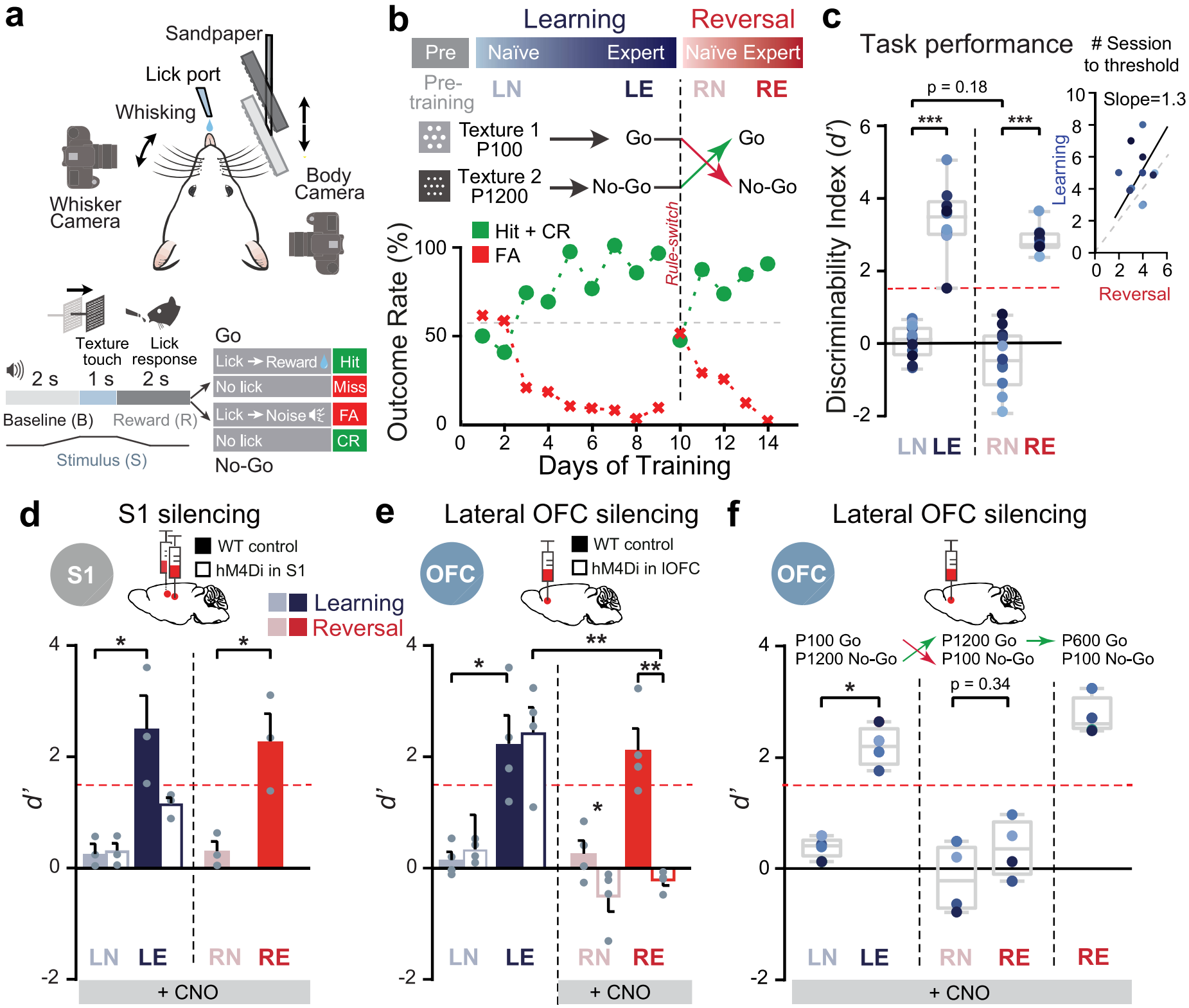
Extended Data Figure 9 | Texture touch-related dynamics in S1 neurons during reversal learning. a, Average Ca^{2+} transient amplitude (ΔF/F) in the stimulus-presentation window for S1 neurons (n = 142 neurons in n = 3 mice, n = 2 sessions/phase). **b,** Scatter plot and histogram comparing texture touch-related selectivity index (*SI*) for the stimulus- presentation window for S1 neurons during learning expert (*LE*) and reversal naïve (*RN*) phase (n = 218 from n = 3 mice, n = 28 sessions). **c,** Scatter plot and histogram comparing *SI* of S1 neurons during *LE* and reversal expert (*RE*) phase (n = 218 neurons from n = 3 mice, n = 28 sessions). **d,** Average Ca^{2+} transient amplitude (ΔF/F) in the stimulus-presentation window for S1 neurons in lOFC silenced mice (n = 87 neurons in n = 2 mice, n = 2 sessions/phase). **e,** Scatter plot and histogram comparing texture touch-related *SI* of S1 neurons during *LE* and *RN* phase in lOFC-silenced mice (n = 165 neurons, n = 25 sessions per phase). **f,** Scatter plot and histogram comparing touch-related *SI* of S1 neurons in lOFC silenced mice during *LE* and *RE* phase (n = 210 neurons in n = 3 mice, n = 28 sessions). **g**, Comparison of *SI* marginal distributions for the three salient periods *LE*, *RN*, and *RE* for lOFC neurons (2D scatter plots not shown), S1 neurons (panels c,d), and S1 neurons in lOFC-silenced mice (panels e,f). All box plots show median, 25^{th} and 75^{th} percentiles as box edges, and 5^{th} and 95^{th} percentiles as whiskers. *p < 0.05, two-sided Wilcoxon rank-sum test.

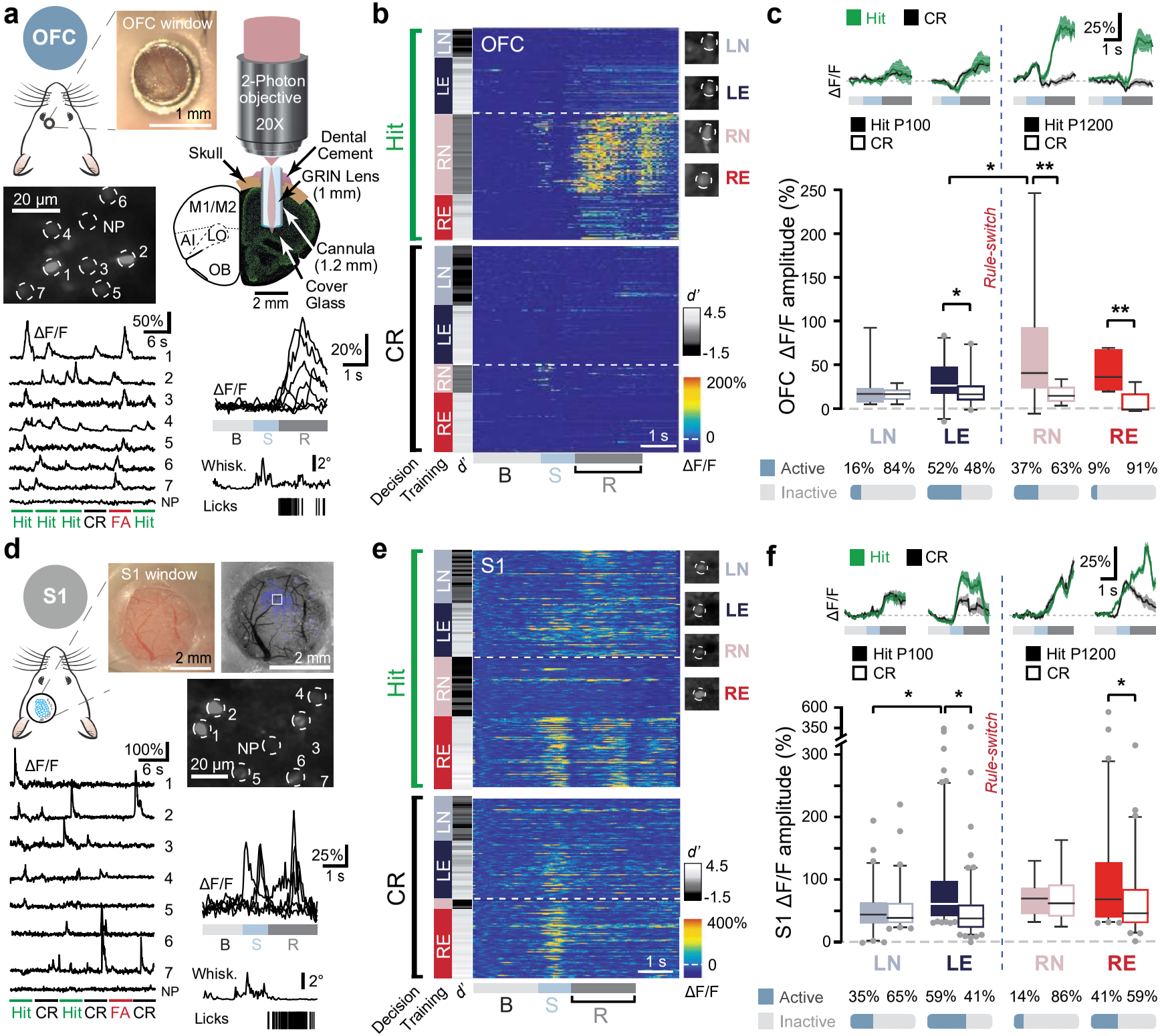
Extended Data Figure 10 | Differential modulation of task variable-relevant events in neuronal responses. a, Schematic diagram of a generalised linear model (GLM, Poisson regression) to predict neural activity from behavioural task variables. Each event was expanded into a series of evenly spaced gaussian filters. **b,** GLM predicting deconvolved neural activity of an example S1 outcome-selective neuron from task variables. **c,** Separate components contributing to the average response of this neuron reveal major sensory modulation together with reward-evoked activity. B-baseline, T-texture touch, R-reward. **d,** To quantify each task variable contribution, the relative fraction of deviance explained is calculated and normalised by the total deviance explained for each neuron both pre- and post-reversal. The reward component in lOFC outcome-selective neurons is significantly greater than the touch related component. **e,** Fraction of deviance explained for each component in separate subsets of S1 neurons reveal distinct modulations for specific task-related events. Notably, responses of outcome selective S1 neuronal responses are mostly explained by reward component. Licking activity seems to modulate S1 neural responses less than reward in each subset. Neurons analysed using GLM are same neurons from Fig. 3. Data is presented as mean ± S.E.M., *p < 0.05, **p < 0.01, two-

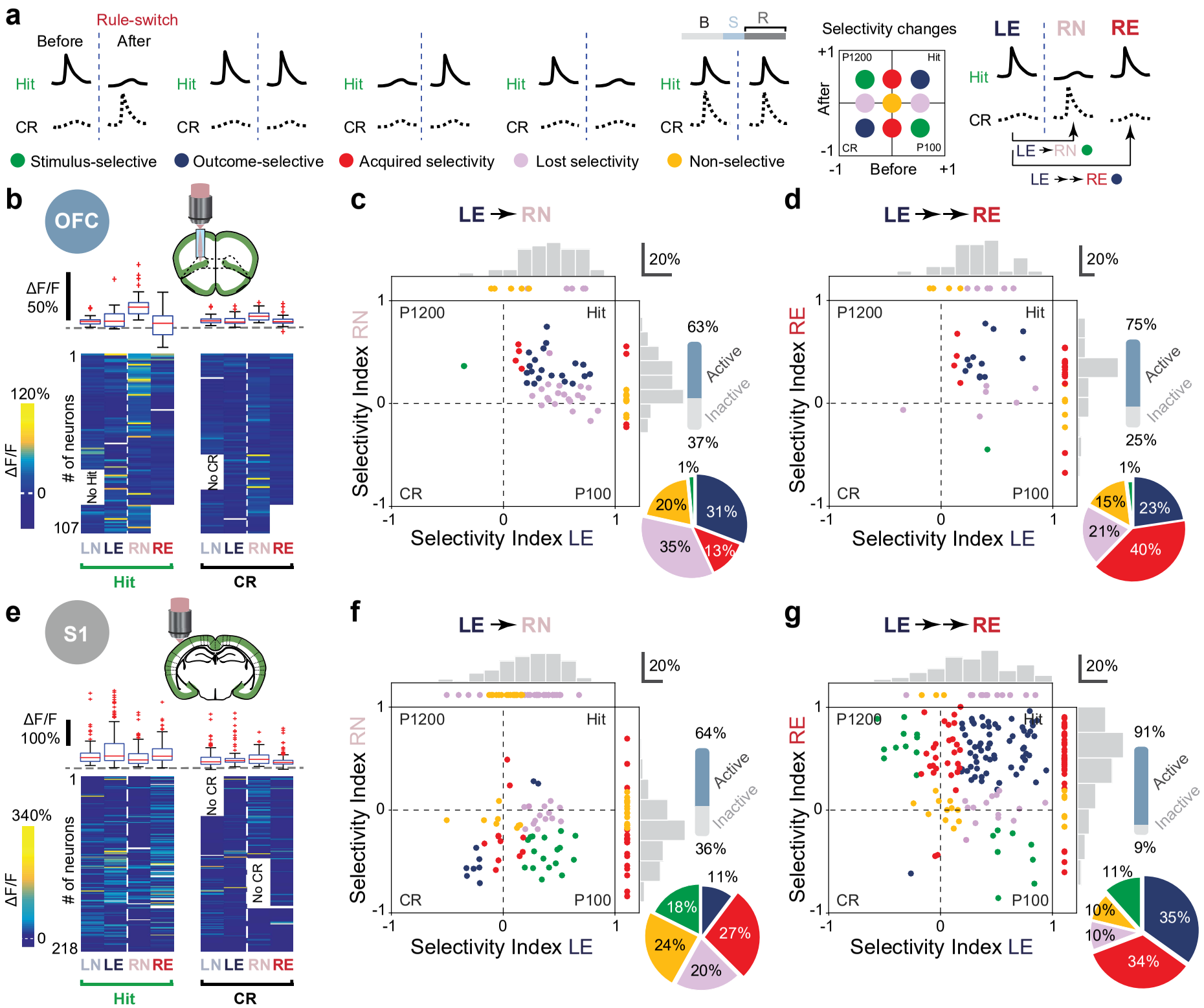
sided Wilcoxon rank-sum test. **f,** Reward-history modulation index (RHMI) for functional subclasses of lOFC neurons and S1 neurons in OFC intact control mice and lOFC-silenced mice (neurons analysed are from Fig. 4b; ns = p > 0.05; bootstrap-permutation test; S.E.M. of RHMI with permutated indices as grey bars).

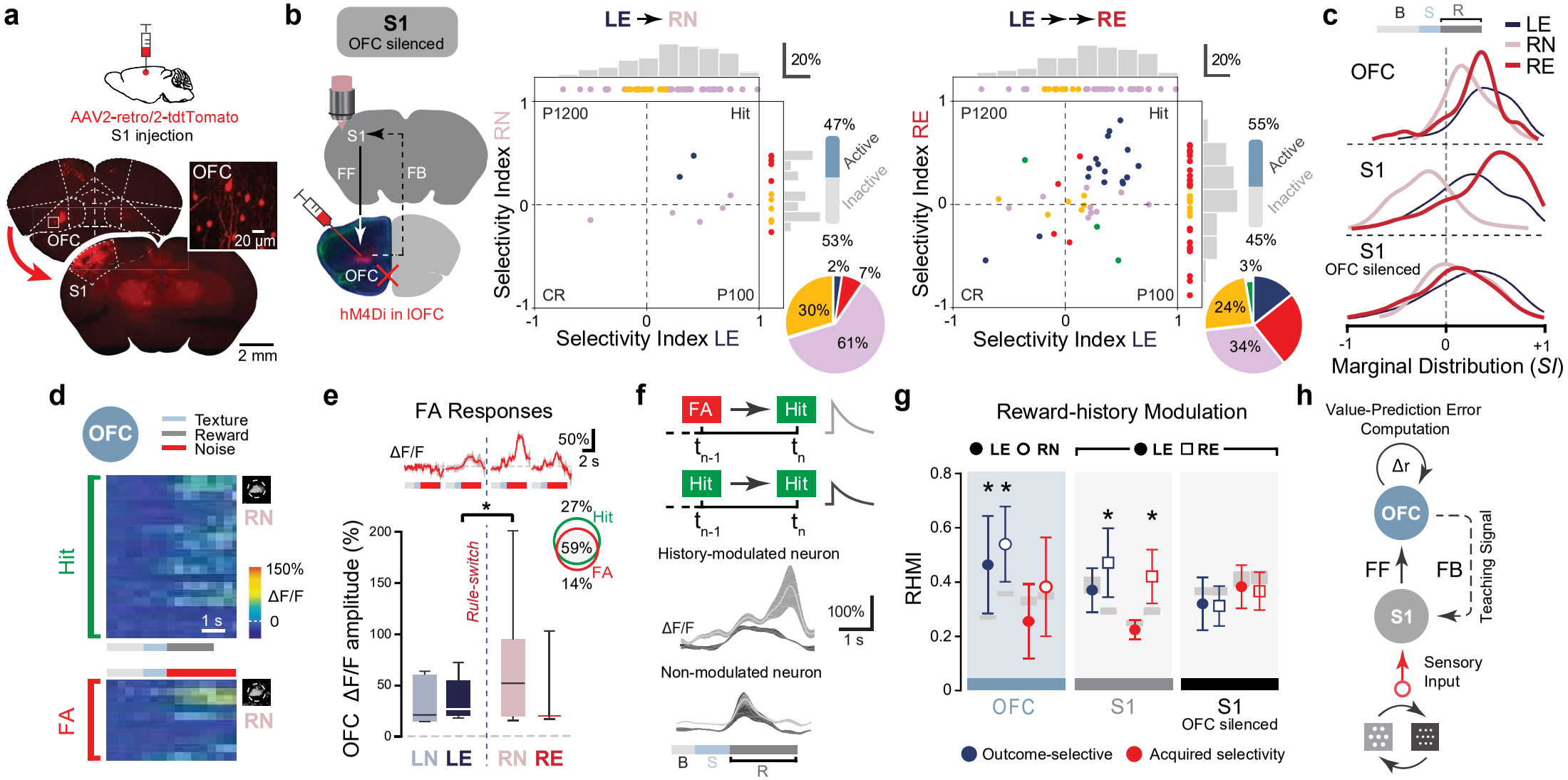
Extended References

31. Mayford, M. et al. Control of memory formation through regulated expression of a CaMKII transgene. *Science (80-.).* **274**, 1678–1683 (1996).
32. Madisen, L. et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
33. Harris, J.A. et al. Anatomical characterization of cre driver mice for neural circuit mapping and manipulation. *Front. Neural Circuits* **8**, 76 (2014).
34. Carandini, M. & Churchland, A.K. Probing perceptual decisions in rodents. *Nat Neurosci* **16**, 824–831 (2013).
35. Bailey, K.R. & Crawley, J.N. Anxiety-related behaviors in mice. *Methods Behav. Anal. Neurosci.* (2009).
36. Farr, T.D., Liu, L., Colwell, K.L., Whishaw, I.Q. & Metz, G.A. Bilateral alteration in stepping pattern after unilateral motor cortex injury: A new test strategy for analysis of skilled limb movements in neurological mouse models. *J. Neurosci. Methods* **153**, 104–13 (2006).
37. Banerjee, A. et al. Jointly reduced inhibition and excitation underlies circuit-wide changes in cortical processing in rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E7287–E7296 (2016).
38. Yang, B. et al. Single-cell phenotyping within transparent intact tissue through whole-body clearing. *Cell* **158**, 945–958 (2014).
39. Chung, K. et al. Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013).
40. Gomez, J.L. et al. Chemogenetics revealed: DREADD occupancy and activation via converted clozapine. *Science (80-.).* **357**, 503–507 (2017).
41. Gilad, A., Gallero-Salas, Y., Groos, D. & Helmchen, F. Behavioral strategy determines frontal or posterior location of short-term memory in neocortex. *Neuron* **99**, 814–828 (2018).
42. Langer, D. et al. HelioScan: A software framework for controlling in vivo microscopy setups with high hardware flexibility, functional diversity and extendibility. *J Neurosci Methods* **215**, 38–52 (2013).
43. Guo, Z. V. et al. Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81**, 179–194 (2014).
44. Sreenivasan, V. et al. Movement initiation signals in mouse whisker motor cortex. *Neuron* **92**, 1368–1382 (2016).
45. Huber, D. et al. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**, 473–478 (2012).

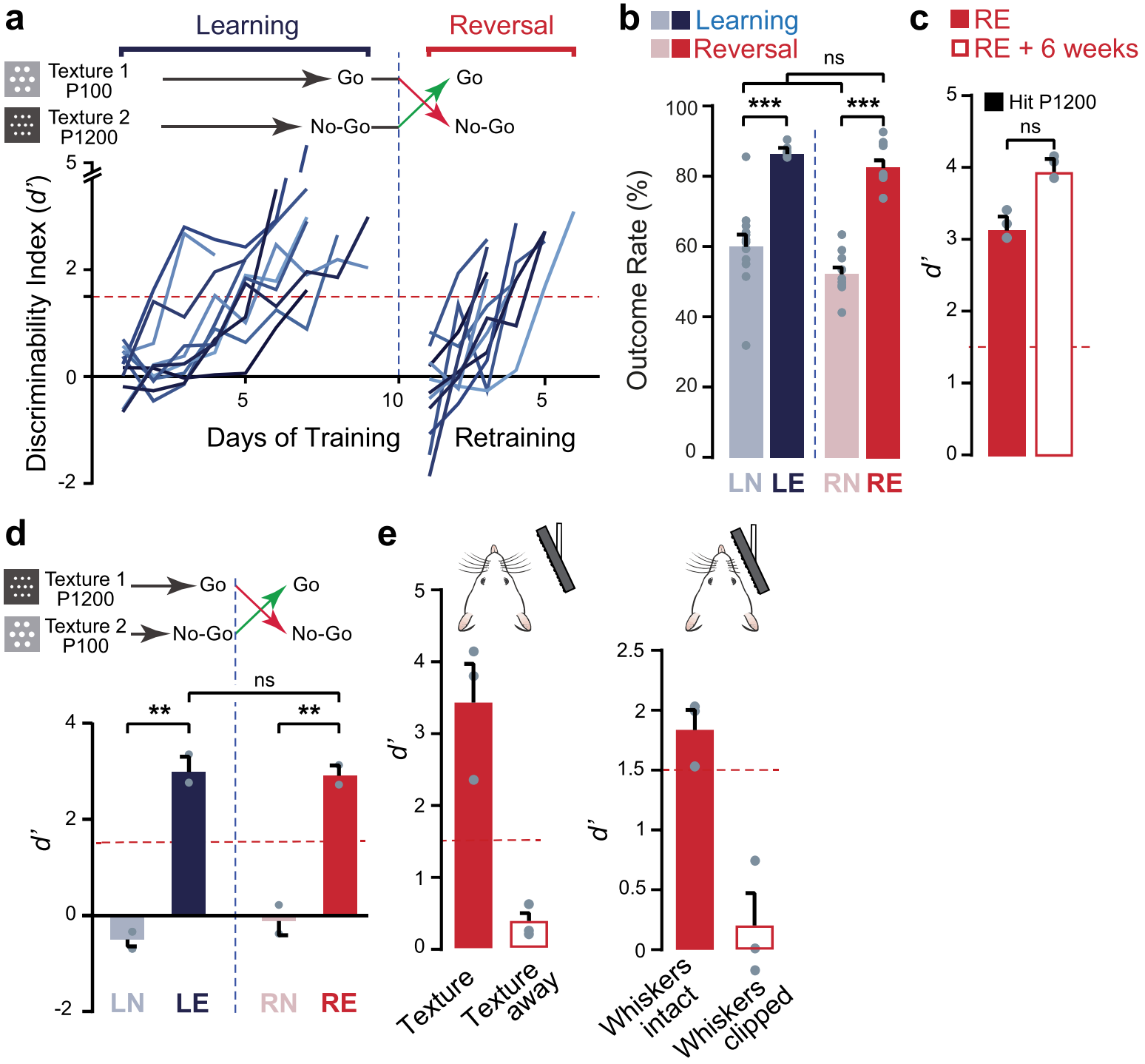

a
Sandpaper
Lick port
Whisking
Body Camera
Whisker Camera
Texture touch
Lick response
2 s
1 s
2 s
Baseline (B)
Reward (R)
Stimulus (S)
Go
Lick → Reward
Hit
No lick
Miss
Lick → Noise
FA
No lick
CR
No-Go
b
Learning
Reversal
Pre
Naïve
Expert
Naïve Expert
Pre-training
LN
LE
RN
RE
Texture 1 P100
Texture 2 P1200
Go
No-Go
Go
No-Go
Hit + CR
FA
Rule-switch
Outcome Rate (%)
Days of Training
c
Task performance
# Session to threshold
p = 0.18
***
***
Discriminability Index (d')
Slope=1.3
Learning
Reversal
LN LE
RN RE
d
S1 silencing
S1
WT control
hM4Di in S1
Learning
Reversal
d'
LN
LE
RN
RE
+ CNO
e
Lateral OFC silencing
OFC
WT control
hM4Di in lOFC
f
Lateral OFC silencing
OFC
P100 Go
P1200 No-Go
P1200 Go
P100 No-Go
P600 Go
P100 No-Go
p = 0.34
+ CNO


a
OFC
OFC window
1 mm
2-Photon objective
20X
Dental Cement
Skull
GRIN Lens (1 mm)
M1/M2
AI
LO
OB
Cannula (1.2 mm)
Cover Glass
2 mm
20 µm
NP
ΔF/F
50%
6 s
20%
1 s
B
S
R
Whisk.
2°
Licks
Hit Hit Hit CR FA Hit
b
OFC
Hit
CR
LN
LE
RN
RE
d'
4.5
-1.5
200%
0
1 s
ΔF/F
Decision
Training
c
Hit
CR
25%
1 s
Hit P100
Hit P1200
OFC ΔF/F amplitude (%)
Rule-switch
Active 16% 84% 52% 48% 37% 63% 9% 91%
Inactive
d
S1
S1 window
2 mm
100%
e
S1
400%
f
S1 ΔF/F amplitude (%)
Active 35% 65% 59% 41% 14% 86% 41% 59%

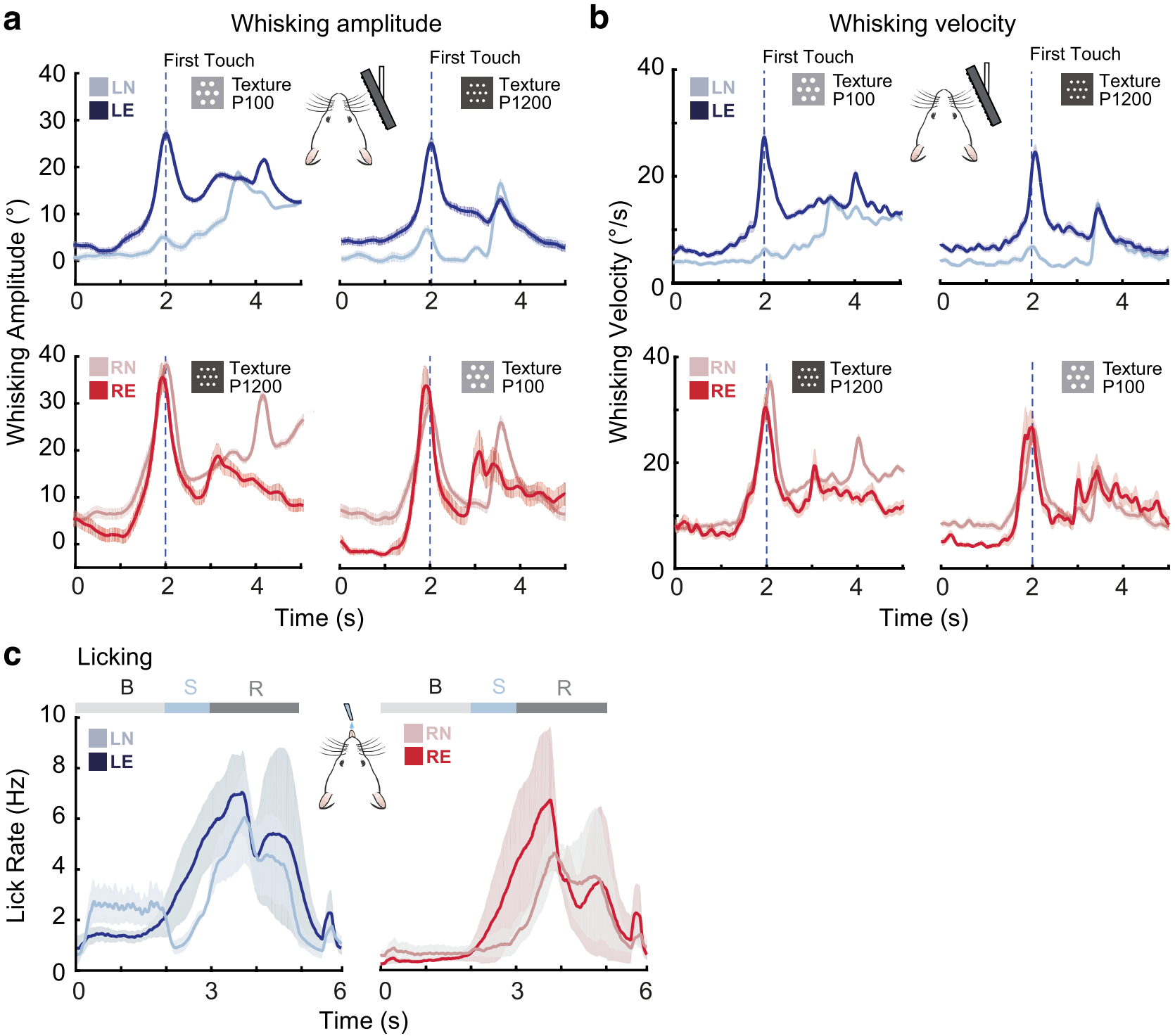

a
Rule-switch
Before
After
Hit
CR
B
S
R
Stimulus-selective
Outcome-selective
Acquired selectivity
Lost selectivity
Non-selective
Selectivity changes
P1200
Hit
CR
P100
After
Before
LE
RN
RE
LE → RN
LE →→ RE
b
OFC
ΔF/F
50%
120%
# of neurons
No Hit
No CR
107
LN LE RN RE
Hit
CR
c
LE → RN
20%
Selectivity Index RN
Selectivity Index LE
63%
Active
Inactive
37%
1%
20%
31%
35%
13%
d
LE →→ RE
Selectivity Index RE
75%
25%
15%
23%
21%
40%
e
S1
ΔF/F
100%
340%
218
f
64%
36%
11%
18%
27%
24%
20%
g
91%
9%
11%
10%
10%
35%
34%


a
AAV2-retro/2-tdtTomato
S1 injection
OFC
20 µm
S1
2 mm
b
S1
OFC silenced
S1
FF
FB
OFC
hM4Di in lOFC
LE → RN
LE →→ RE
20%
P1200
Hit
CR
P100
Selectivity Index RN
Selectivity Index RE
Selectivity Index LE
47%
Active
Inactive
53%
2%
7%
30%
61%
55%
45%
3%
24%
34%
c
B
S
R
LE
RN
RE
OFC
S1
S1
OFC silenced
Marginal Distribution (SI)
d
OFC
Texture
Reward
Noise
Hit
RN
150%
ΔF/F
0
1 s
FA
RN
e
FA Responses
ΔF/F
50%
2 s
OFC ΔF/F amplitude (%)
Rule-switch
27%
Hit
59%
FA
14%
LN LE
RN RE
f
FA
Hit
Hit
Hit
History-modulated neuron
ΔF/F
100%
1 s
Non-modulated neuron
B
S
R
g
Reward-history Modulation
LE RN
LE RE
RHMI
OFC
S1
S1
OFC silenced
Outcome-selective
Acquired selectivity
h
Value-Prediction Error
Computation
Δr
OFC
FF
FB
Teaching Signal
S1
Sensory
Input

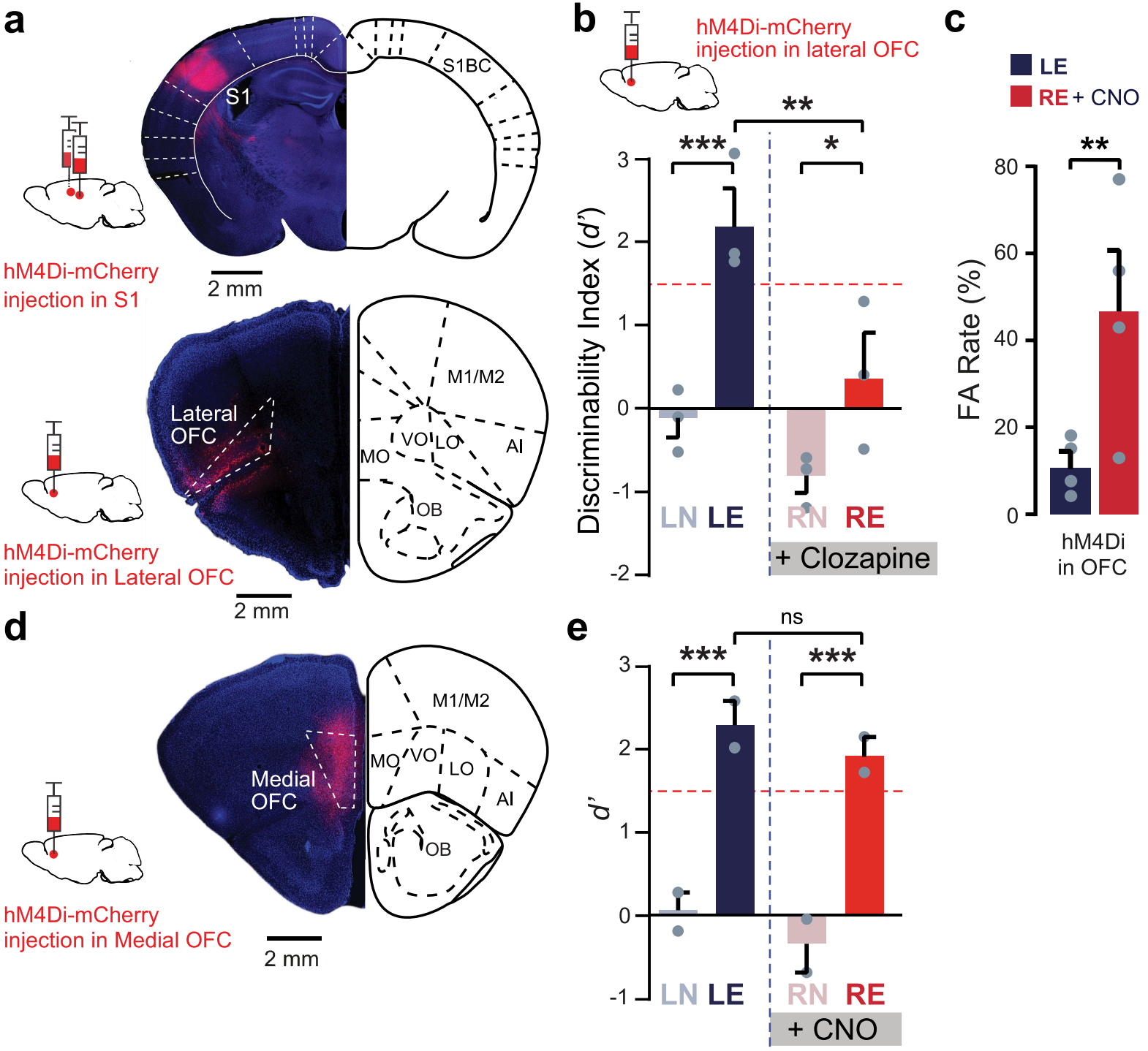
Suppl. Fig. 1


a
Learning
Reversal
Texture 1
P100
Texture 2
P1200
Go
No-Go
Go
No-Go
Discriminability Index (d')
Days of Training
Retraining
b
Learning
Reversal
ns
***
***
Outcome Rate (%)
LN LE RN RE
c
RE
RE + 6 weeks
Hit P1200
ns
d'
d
Texture 1
P1200
Texture 2
P100
Go
No-Go
Go
No-Go
ns
**
**
d'
LN LE RN RE
e
d'
Texture
Texture away
Whiskers intact
Whiskers clipped

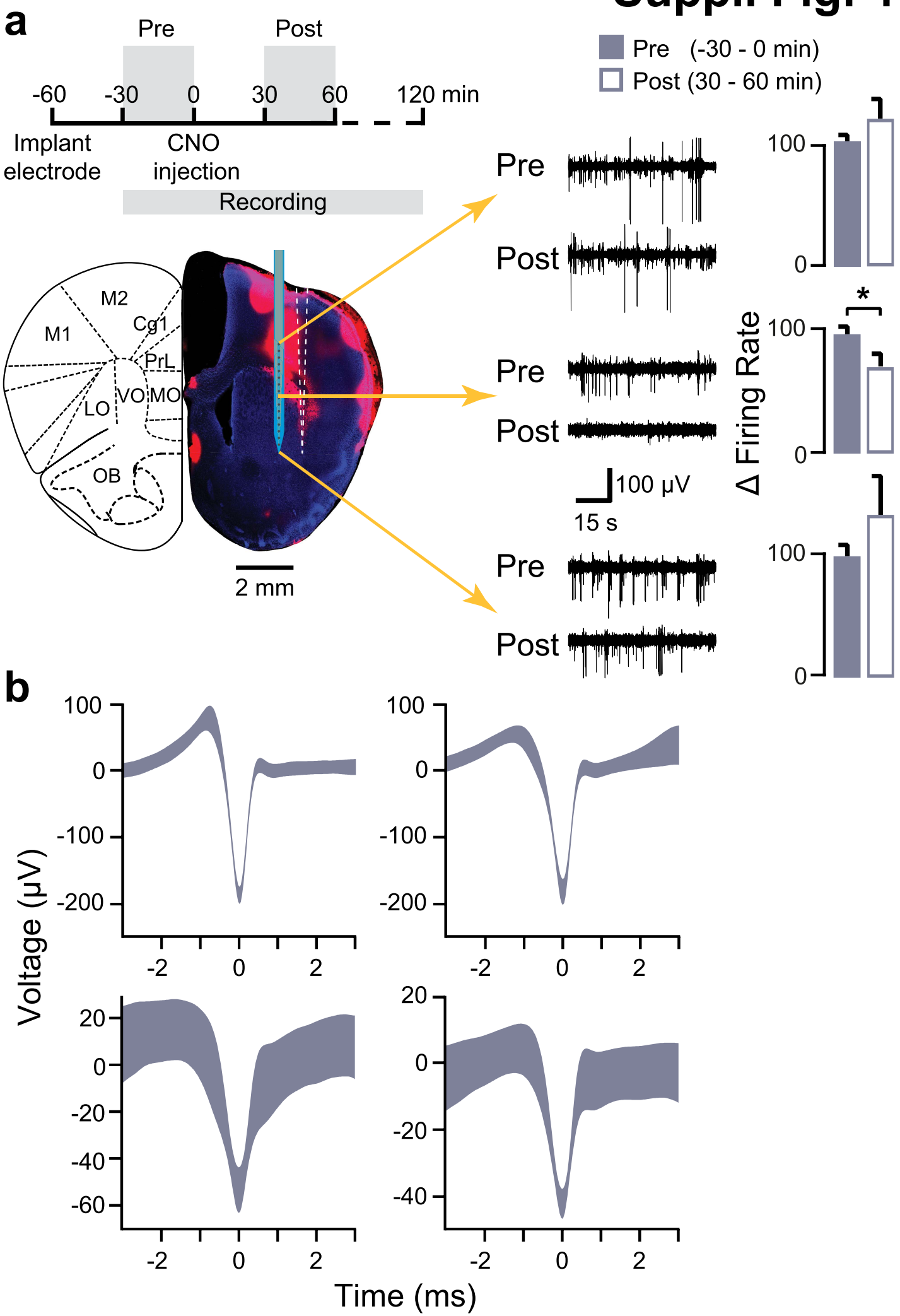
Suppl. Fig. 2


a
Whisking amplitude
First Touch
First Touch
LN
LE
Texture
P100
Texture
P1200
RN
RE
Texture
P1200
Texture
P100
Whisking Amplitude (°)
Time (s)
b
Whisking velocity
First Touch
First Touch
LN
LE
Texture
P100
Texture
P1200
RN
RE
Texture
P1200
Texture
P100
Whisking Velocity (°/s)
Time (s)
c
Licking
B
S
R
B
S
R
LN
LE
RN
RE
Lick Rate (Hz)
Time (s)

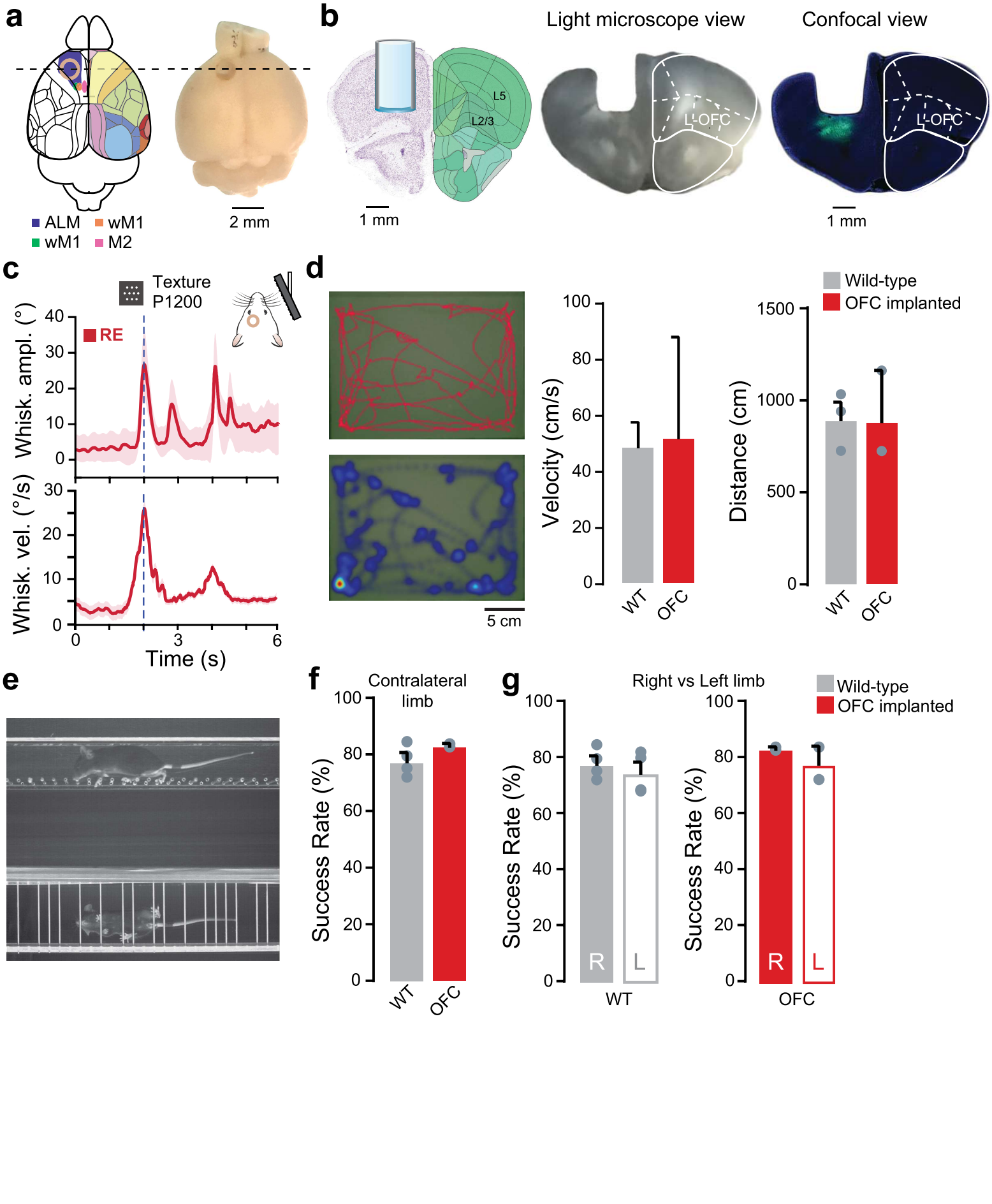
Suppl. Fig. 3


a
S1
S1BC
hM4Di-mCherry
injection in S1
2 mm
Lateral
OFC
M1/M2
VO
LO
AI
MO
OB
hM4Di-mCherry
injection in Lateral OFC
2 mm
b
hM4Di-mCherry
injection in lateral OFC
**
***
*
Discriminability Index (d')
3
2
1
0
-1
-2
LN LE
RN RE
+ Clozapine
c
LE
RE + CNO
**
FA Rate (%)
80
60
40
20
0
hM4Di
in OFC
d
Medial
OFC
M1/M2
MO
VO
LO
AI
OB
hM4Di-mCherry
injection in Medial OFC
2 mm
e
ns
***
***
d'
3
2
1
0
-1
LN LE
RN RE
+ CNO

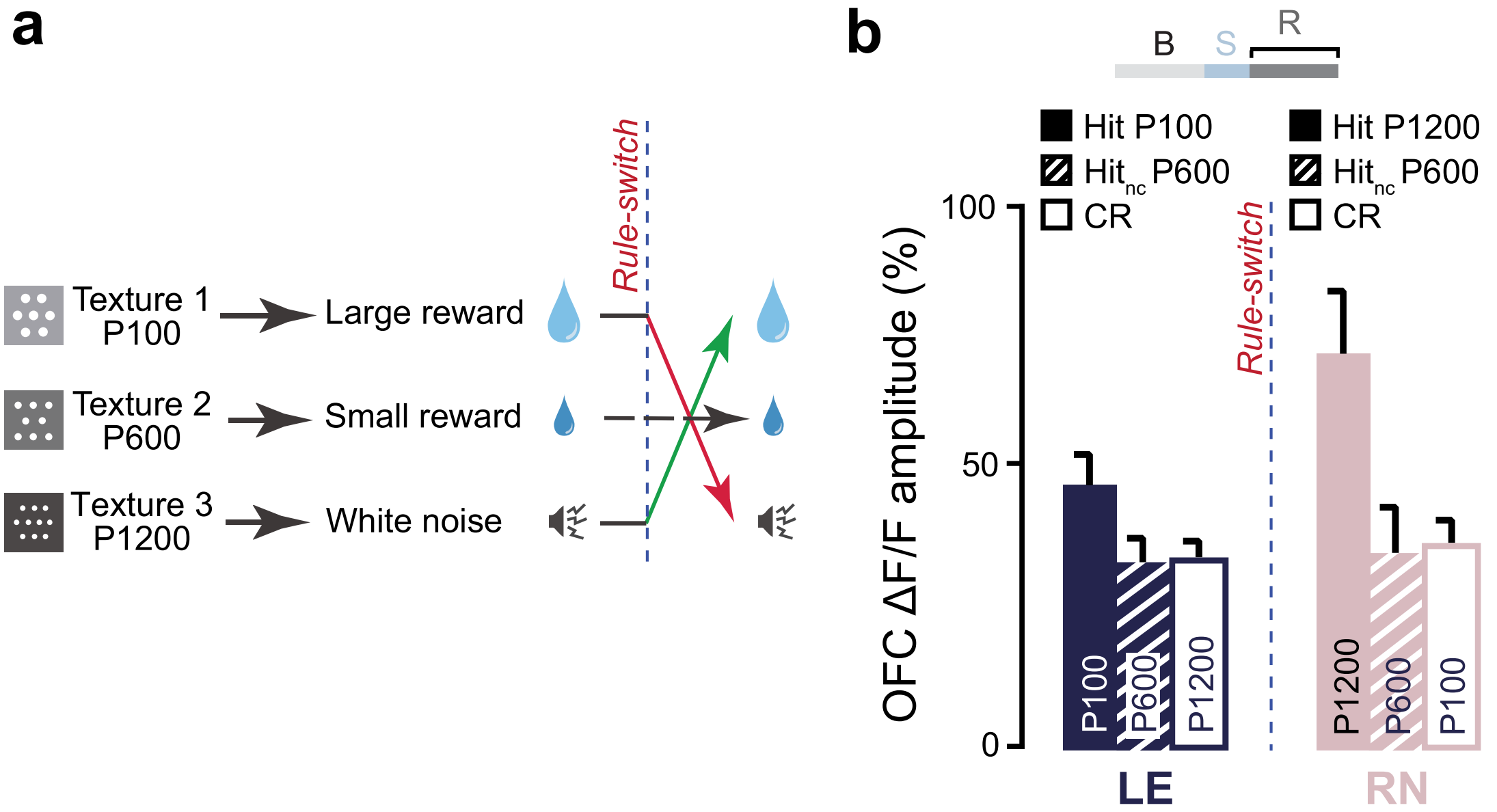
Suppl. Fig. 4


a
Pre
Post
-60
-30
0
30
60
120 min
Implant electrode
CNO injection
Recording
M2
M1
Cg1
PrL
VO
MO
LO
OB
2 mm
Pre
Post
100 µV
15 s
Pre (-30 - 0 min)
Post (30 - 60 min)
Δ Firing Rate
100
0
*
b
Voltage (µV)
Time (ms)

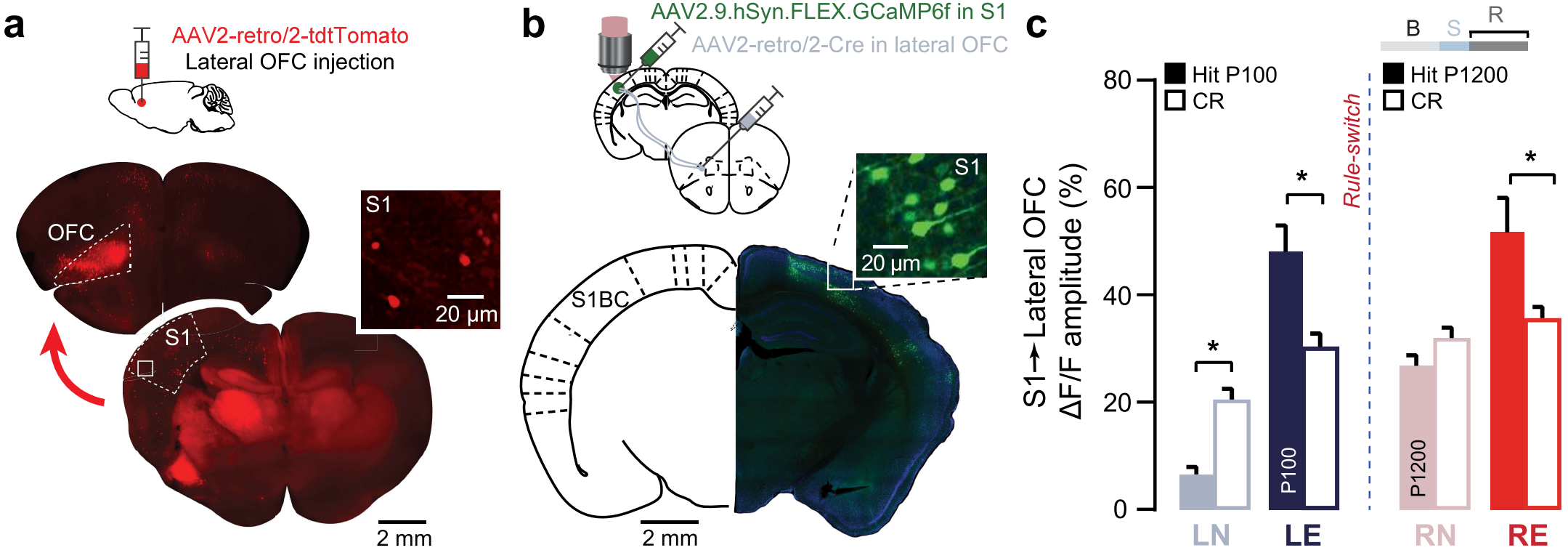
Suppl. Fig. 5


a
ALM
wM1
wM1
M2
2 mm
b
L5
L2/3
1 mm
Light microscope view
Confocal view
L-OFC
L-OFC
1 mm
c
Texture
P1200
RE
Whisk. ampl. (°)
Whisk. vel. (°/s)
Time (s)
d
5 cm
Velocity (cm/s)
Distance (cm)
Wild-type
OFC implanted
WT
OFC
e
f
Contralateral limb
Success Rate (%)
WT
OFC
g
Right vs Left limb
Success Rate (%)
R
L
WT
OFC

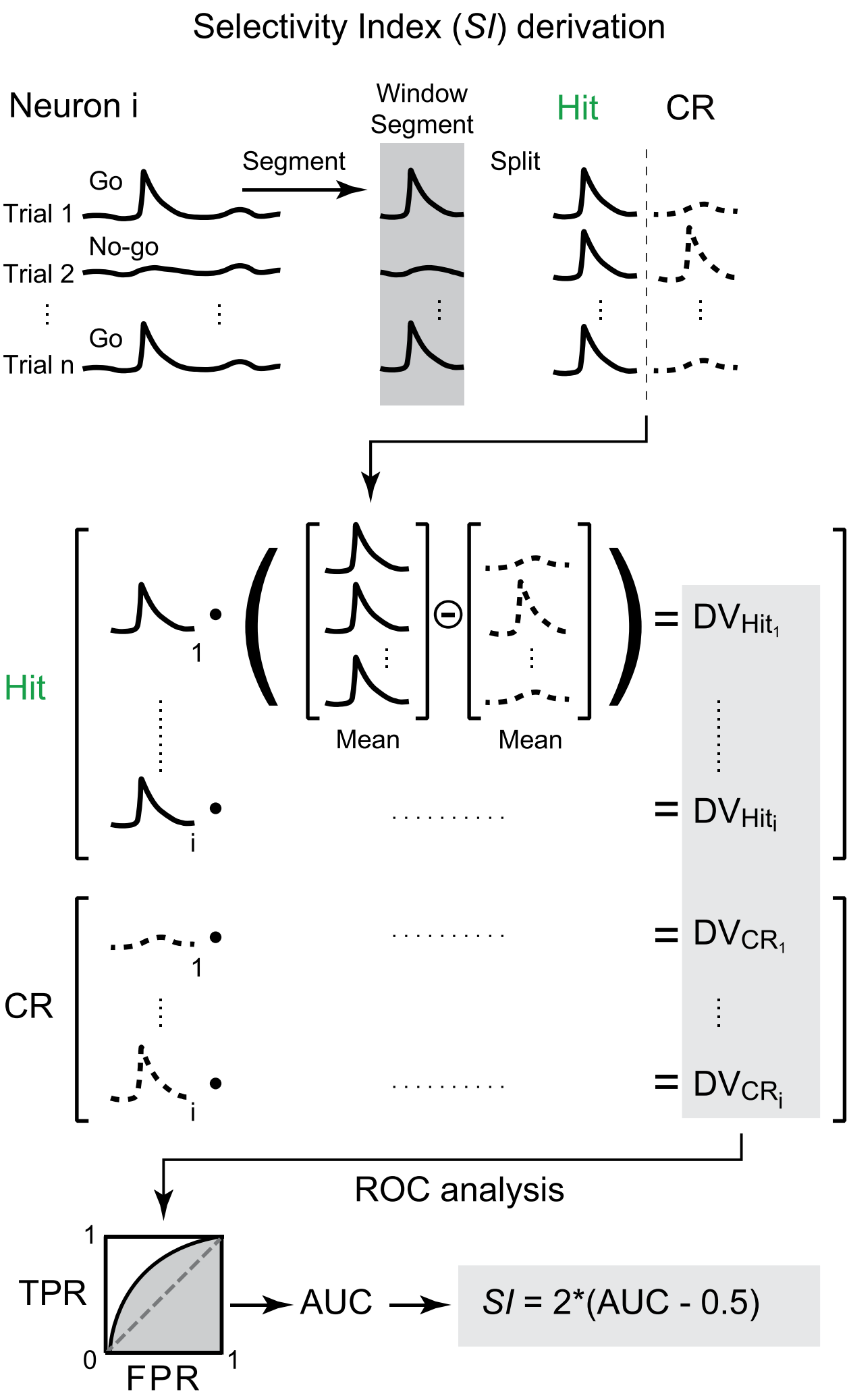
Suppl. Fig. 6


a
Texture 1
P100
Large reward
Texture 2
P600
Small reward
Texture 3
P1200
White noise
Rule-switch
b
B
S
R
Hit P100
Hit$_{nc}$ P600
CR
Hit P1200
Hit$_{nc}$ P600
CR
OFC ΔF/F amplitude (%)
100
50
0
Rule-switch
P100
P600
P1200
P1200
P600
P100
LE
RN

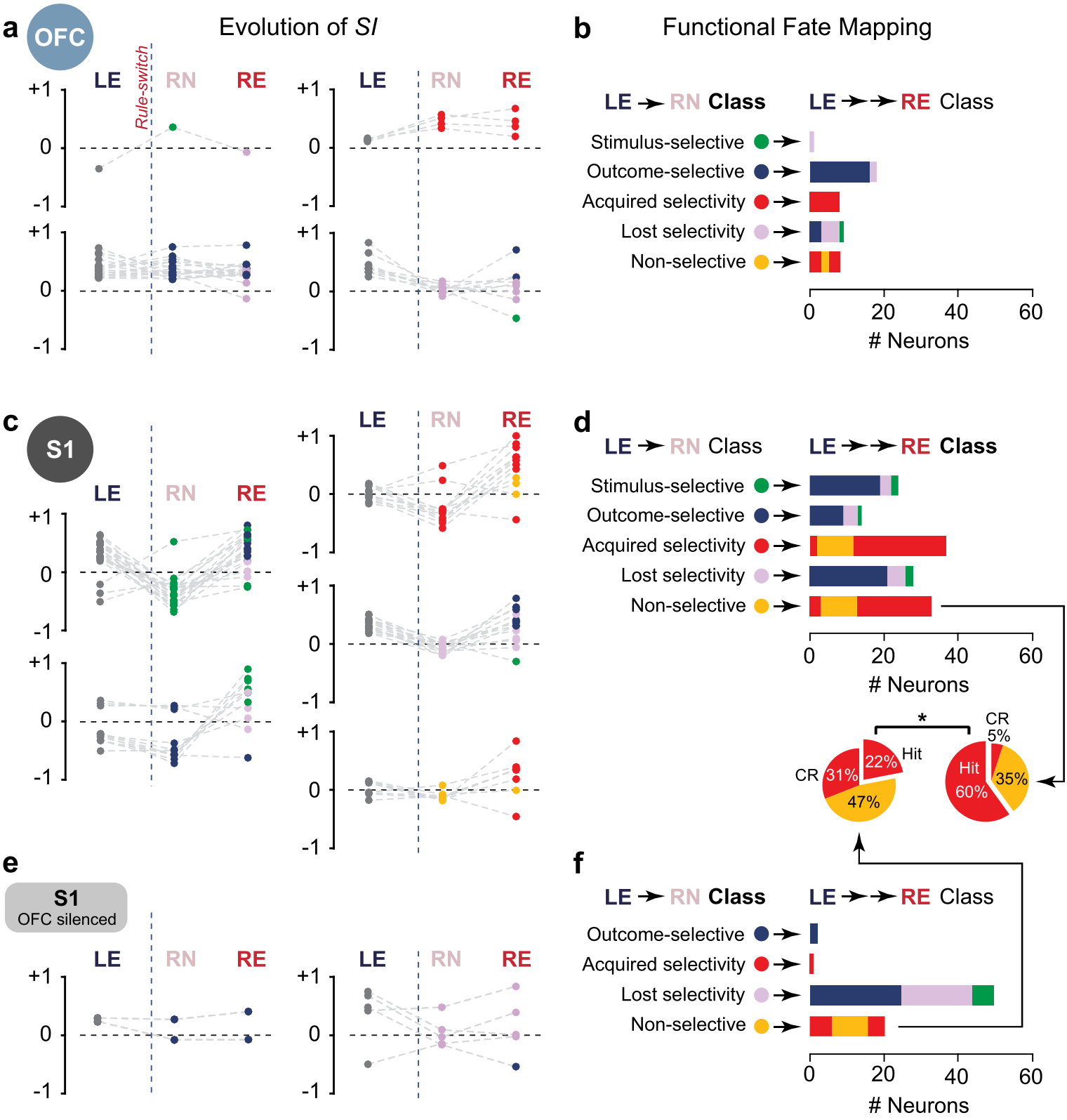
Suppl. Fig. 7


a
AAV2-retro/2-tdtTomato
Lateral OFC injection
OFC
S1
S1
20 µm
2 mm
b
AAV2.9.hSyn.FLEX.GCaMP6f in S1
AAV2-retro/2-Cre in lateral OFC
S1
20 µm
S1BC
2 mm
c
B
S
R
Hit P100
CR
Hit P1200
CR
Rule-switch
S1→Lateral OFC
ΔF/F amplitude (%)
80
60
40
20
0
P100
P1200
LN
LE
RN
RE
*
*
*

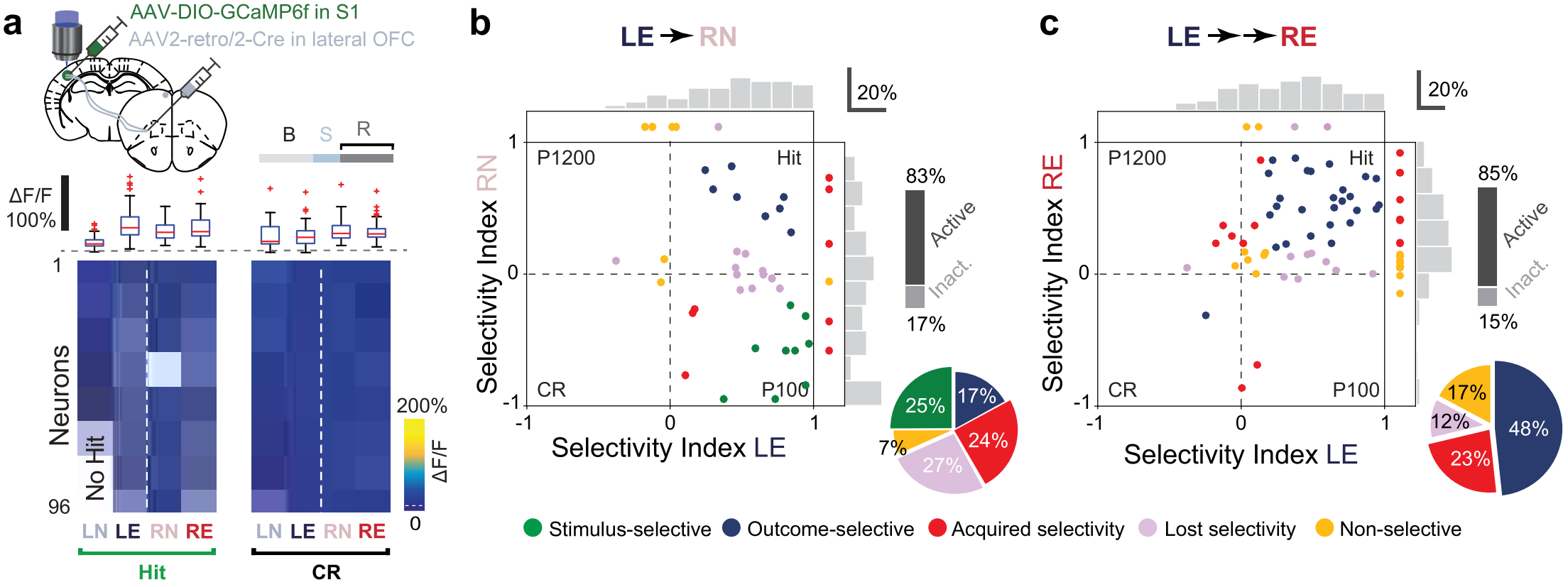
Suppl. Fig. 8


Selectivity Index (SI) derivation
Neuron i
Window Segment
Hit
CR
Segment
Split
Go
Trial 1
No-go
Trial 2
Go
Trial n
Hit
1
i
Mean
Mean
= DV Hit 1
= DV Hit i
CR
1
i
= DV CR 1
= DV CR i
ROC analysis
1
TPR
0
FPR
1
AUC
SI = 2*(AUC - 0.5)

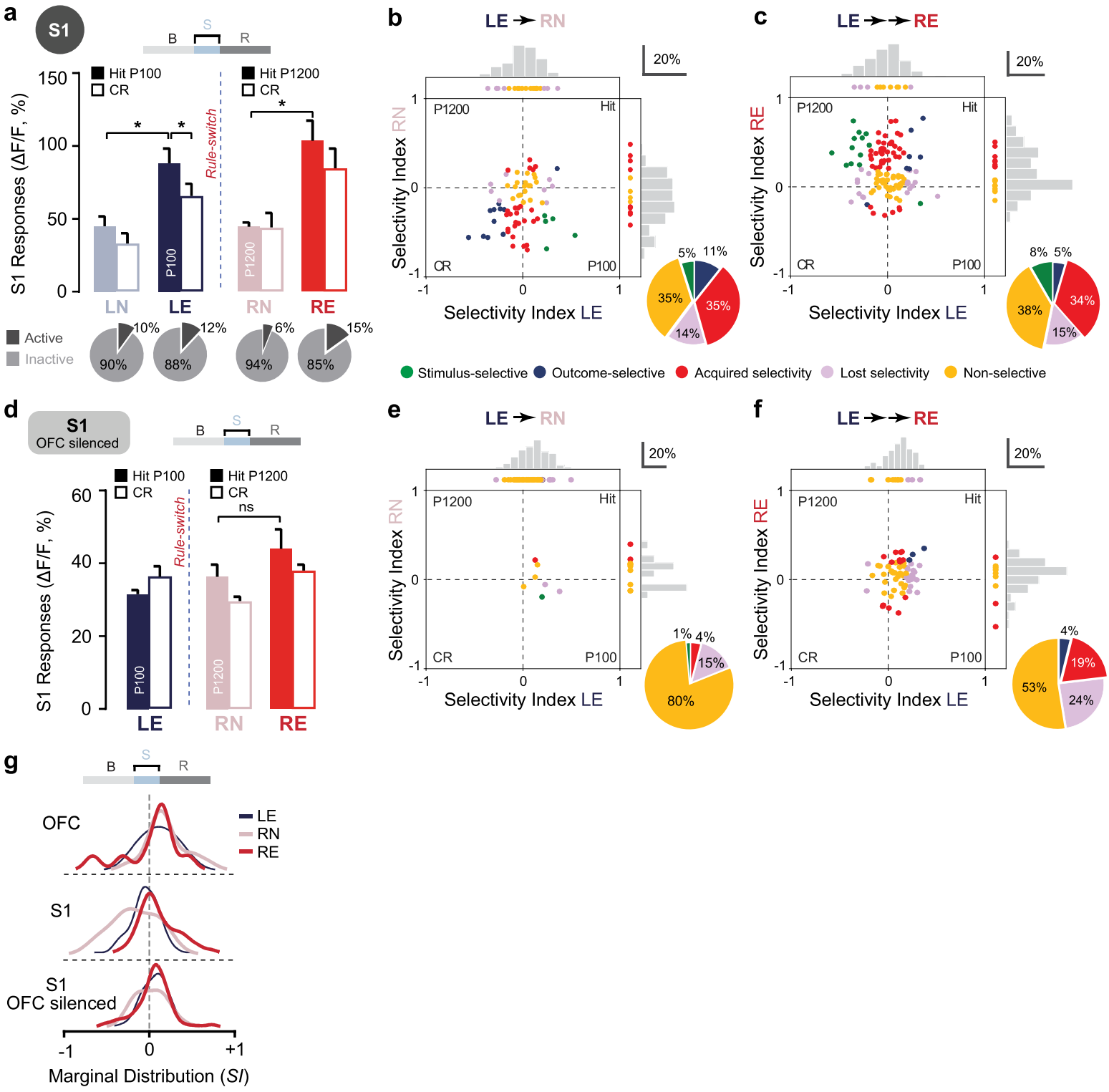
Suppl. Fig. 9


a
OFC
Evolution of SI
Rule-switch
LE
RN
RE
+1
0
-1
b
Functional Fate Mapping
LE → RN Class
LE →→ RE Class
Stimulus-selective
Outcome-selective
Acquired selectivity
Lost selectivity
Non-selective
0
20
40
60
Neurons
c
S1
d
*
CR
5%
Hit
22%
CR
31%
47%
Hit
60%
35%
e
S1
OFC silenced
f

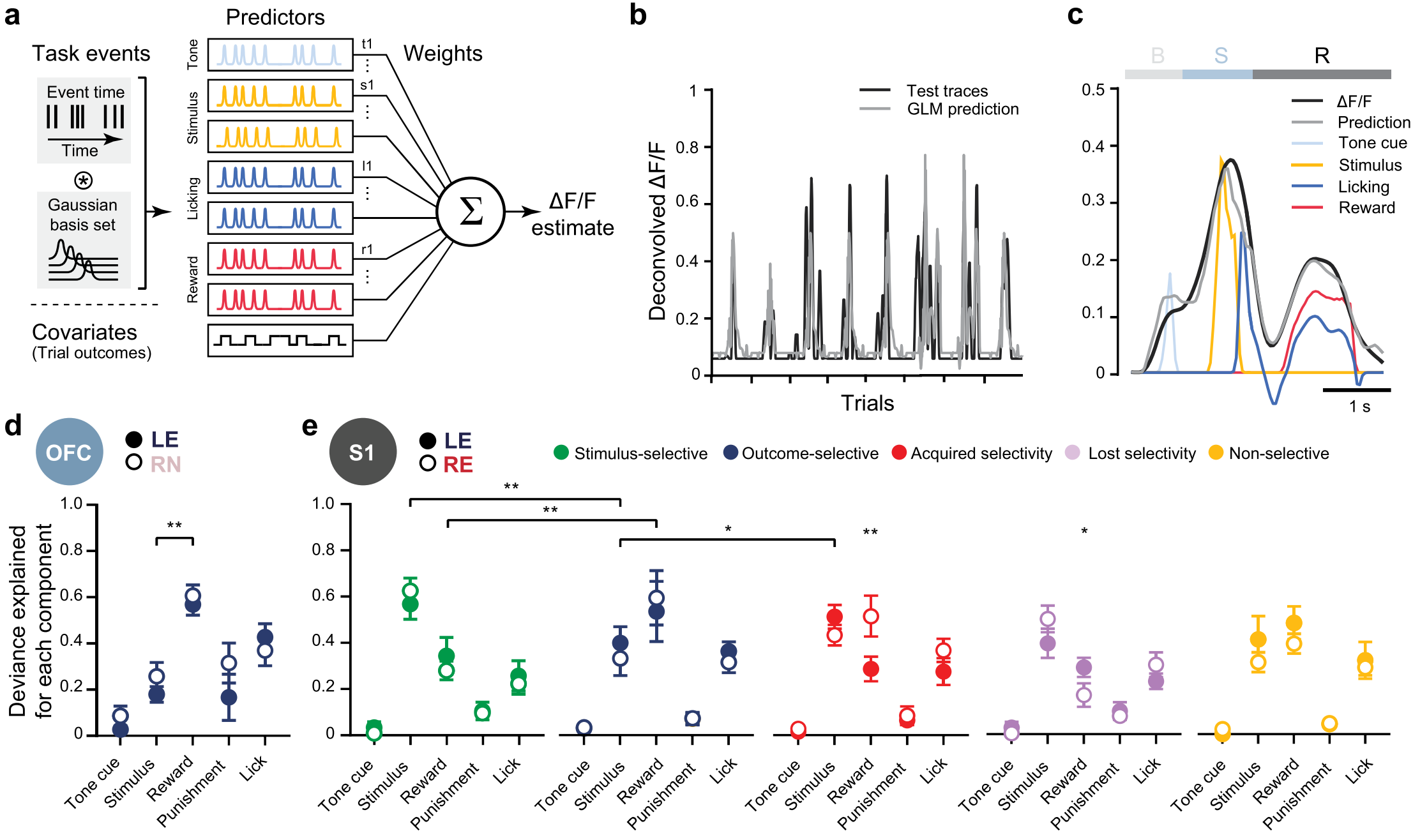
Suppl. Fig. 10


a
AAV-DIO-GCaMP6f in S1
AAV2-retro/2-Cre in lateral OFC
B S R
ΔF/F
100%
1
Neurons
No Hit
96
LN LE RN RE
LN LE RN RE
Hit
CR
200%
ΔF/F
0
b
LE → RN
20%
P1200
Hit
CR
P100
Selectivity Index RN
Selectivity Index LE
1
0
-1
83%
Active
Inact.
17%
25%
17%
24%
7%
27%
c
LE → → RE
20%
P1200
Hit
CR
P100
Selectivity Index RE
Selectivity Index LE
85%
Active
Inact.
15%
17%
12%
48%
23%
Stimulus-selective
Outcome-selective
Acquired selectivity
Lost selectivity
Non-selective

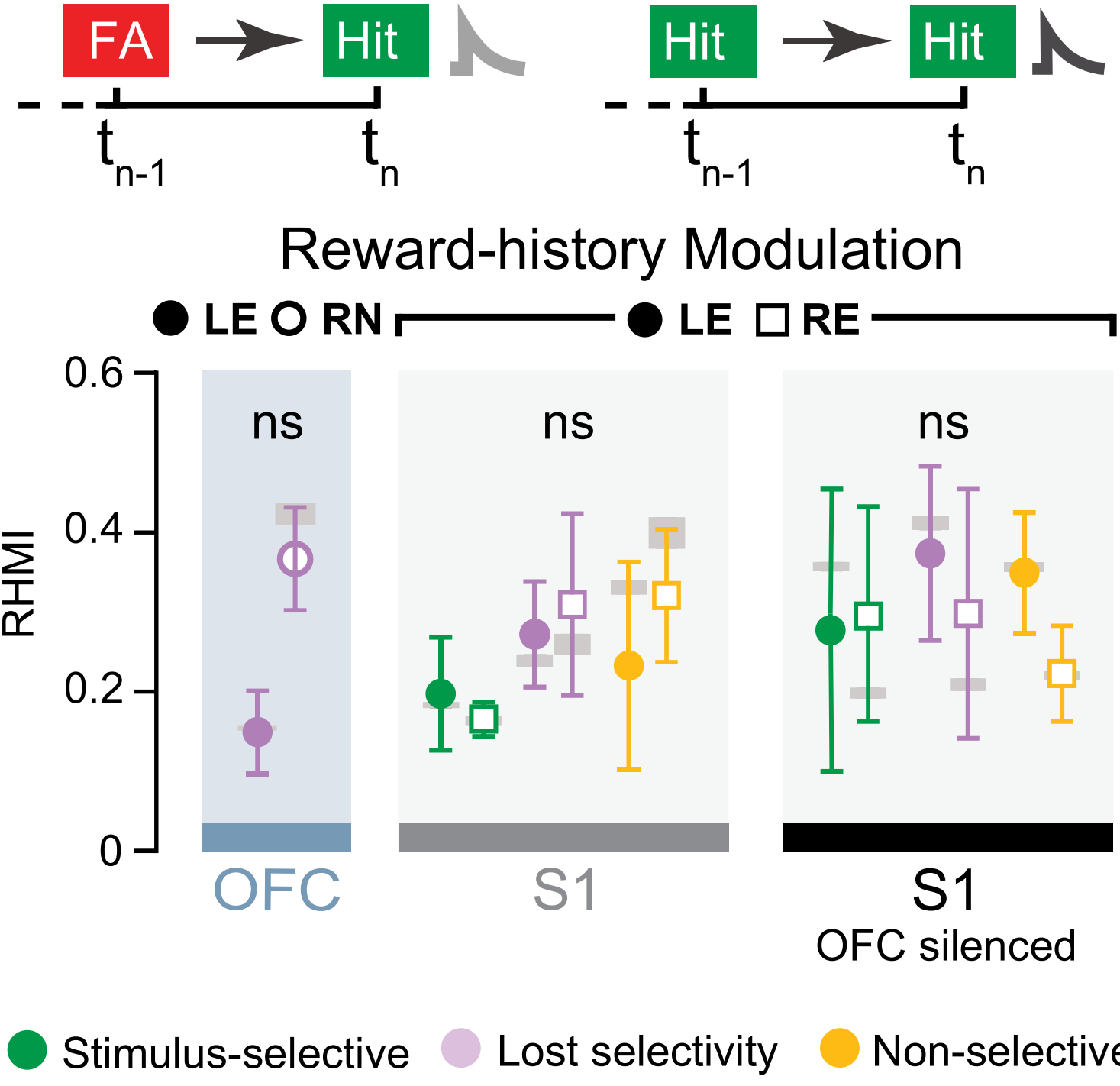
Suppl. Fig. 11


a
S1
B
S
R
Hit P100
CR
Hit P1200
CR
Rule-switch
S1 Responses (ΔF/F, %)
150
100
50
0
P100
P1200
LN
LE
RN
RE
Active
Inactive
10%
90%
12%
88%
6%
94%
15%
85%
b
LE → RN
20%
Selectivity Index RN
Selectivity Index LE
P1200
Hit
CR
P100
5%
11%
35%
35%
14%
c
LE →→ RE
20%
Selectivity Index RE
Selectivity Index LE
8%
5%
38%
34%
15%
Stimulus-selective
Outcome-selective
Acquired selectivity
Lost selectivity
Non-selective
d
S1
OFC silenced
ns
e
1%
4%
15%
80%
f
4%
19%
53%
24%
g
OFC
S1
S1
OFC silenced
-1
0
+1
Marginal Distribution (SI)

Suppl. Fig. 12


a
Predictors
Task events
Event time
Time
Gaussian basis set
Covariates
(Trial outcomes)
Tone
Stimulus
Licking
Reward
t1
s1
l1
r1
Weights
ΔF/F estimate
b
Test traces
GLM prediction
Deconvolved ΔF/F
Trials
c
B
S
R
ΔF/F
Prediction
Tone cue
Stimulus
Licking
Reward
1 s
d
OFC
LE
RN
Deviance explained for each component
Tone cue
Stimulus
Reward
Punishment
Lick
e
S1
LE
RE
Stimulus-selective
Outcome-selective
Acquired selectivity
Lost selectivity
Non-selective

Suppl. Fig. 13


FA
Hit
Hit
Hit
t_{n-1}
t_n
t_{n-1}
t_n
Reward-history Modulation
LE RN
LE RE
RHMI
0.6
0.4
0.2
0
ns
ns
ns
OFC
S1
S1
OFC silenced
Stimulus-selective
Lost selectivity
Non-selective